Keeping it Real: Revisiting a Real-Space Approach to Running Ensembles of Cosmological N-body Simulations

Chris Orban[1,2*]

(1) Center for Cosmology and Astro-Particle Physics, The Ohio State University, 191 W Woodruff Ave, Columbus, OH 43210
(2) Department of Physics, The Ohio State University, 191 W Woodruff Ave, Columbus, OH 43210

In setting up initial conditions for cosmological N-body simulations there are, fundamentally, two choices: either maximizing the correspondence of the initial density field to the assumed fourier-space clustering or, instead, matching to the real-space clustering. As a stringent test of both approaches, I perform ensembles of simulations using power law models and exploit the self-similarity of these initial conditions to quantify the accuracy of the results. Originally proposed by Pen 1997 and implemented by Sirko 2005, I show that the real-space motivated approach, which allows the DC mode to vary, performs well in exhibiting the expected self-similar behavior in the mean $\xi(r)$ and $P(k)$ and in both methods this behavior extends below the scale of the initial mean interparticle spacing. I also test the real-space method with simulations of a simplified, powerlaw model for baryon acoustic oscillations, again with success, and mindful of the need to generate mock catalogs using simulations I show extensive (powerlaw) tests for the halo mass function and halo bias in my simulations. Although requiring a few to many times more simulations than the standard, fourier-space method to reach a given certainty on the correlation function or power spectrum, I find that the real-space method is more reliable for modeling $P(k)$ when the clustering level becomes significant on the scale of the simulation box. However, even for the niche application of simulating extremely red power spectra ($n_{\rm eff} \lesssim -2$) or small box ΛCDM simulations the ξ-sampled approach in its original form is not especially well-suited for this regime. I conclude by discussing some possibilities for optimizing or hybridizing the real-space method for more general use and an appendix demonstrates the potential for using perturbation theory to model the effect of the box scale on the simulated growth of structure in conventional P-sampled simulations.

I. INTRODUCTION

Next-generation astronomical surveys will demand increasingly precise predictions from theory in order to properly interpret observations and constrain the nature of dark energy. As emphasized by [1], this will be a challenging task: inaccuracies in the predictions of halo abundance and halo bias, for example, can affect cosmological inferences [2], and measurements of the baryon acoustic oscillations (BAO) clustering feature will eventually reach the stage where theoretical estimates of the shift of this feature from non-linear dynamics become important [3]. Although current state-of-the-art cosmological N-body simulations, given a specific set of cosmological parameters, are in many ways well-equipped to deliver highly precise predictions of the dark matter two-point correlation function and power spectrum for a relatively wide range of scales [4], the difficult-to-estimate covariances of these statistics are also crucial for placing constraints on cosmological parameters [5, 6]. To stay ahead of the observational demands, any methodological improvement to the current framework of setting up and running cosmological N-body simulations is therefore welcome, even if only relevant to certain applications and in certain regimes.

In this paper I explore the predictions of a method for running ensembles of simulations that, unlike the conventional method, is designed to maximize correspondence between simulated real-space clustering statistics (e.g. $\sigma_8, \xi(r)$) and the real-space properties of the assumed cosmological model. Originally proposed by Pen [7] and implemented by Sirko [8][1], this method allows the DC mode of each simulation (in an ensemble of simulations) to vary according to the clustering power on the scale of the box in much the same way that the density within randomly placed boxes in the real universe will fluctuate around the mean density. Although it is now common practice, using the conventional method, to set the DC mode in each simulation to zero, in the early days of fully cosmological N-body simulations this was not always done [e.g. 9]. This issue has also been discussed in the context of artificially changing the DC mode of an existing simulation as a way of scaling a simulation completed with a certain set of cosmological parameters to a slightly different model [10–13].

In the Sirko [8] framework the initial power spectrum used with the Zeldovich [14] (and by extension 2LPT [15, 16]) approximation is convolved such that the matter correlation function matches exactly the linear theory correlation function for $r < L_{\rm box}/2$, while for $r > L_{\rm box}/2$ the correlation function is set to zero. With this in mind

* orban@physics.osu.edu

[1] Once publicly available, the code can still be obtained through http://web.archive.org

Sirko refers to this approach as "ξ-sampled" initial conditions (ICs), while the standard method is referred to as "P-sampled", since by using an unconvolved linear theory power spectrum with the Zeldovich approximation the initial conditions are instead matched to the fourier space clustering statistics. The ξ-sampled strategy, by matching the correlation function out to $r = L_{\rm box}/2$, should avoid biases on all real space statistics, since the rms overdensity in spheres, $\sigma(R)$, is simply related to the correlation function, and the halo mass function to good approximation is only a function of $\sigma(R)$. Without this convolution these real space statistics become biased (i.e. from $P(k) = 0$ for $k \lesssim 2\pi/L_{\rm box}$), as discussed by [7] and [17].

Although a number of groups have published results using the initial conditions code developed by Sirko, which was the among the first include the 2nd order Lagrangian corrections [15, 16] to the Zeldovich [14] displacements, the code is very seldom used to generate ξ-sampled ICs. To my knowledge, only Reid et al. [18] have utilized the code in this mode, citing the success of convergence tests in [19]. In that study they create mock catalogues from a suite of 42 simulations with $L_{\rm box} = 558h^{-1}$ Mpc, and $N = 512^3$ for comparison with SDSS LRG data [20]. They chose the ξ-sampled method for this task, citing the attractive feature of allowing the DC mode of the box to vary, thereby modeling the power spectrum covariance of real surveys more realistically. [19] and Appendix A of [18] present a wide variety of convergence tests that explore the effects of increasing the resolution with either fixed initial conditions (i.e. with a particular randomly sampled value for the DC mode) or for a set of a few initial conditions realizations. More recently, [21] argued that the DC mode should be re-introduced and compared the results of five $L_{\rm box} = 20h^{-1}$ Mpc, ΛCDM simulations using the ξ-sampled method to a high-resolution, standard-method simulation with $L_{\rm box} = 80h^{-1}$ Mpc, finding good correspondence between the results for the variance of the halo mass function.

This study systematically explores the predictions of the two different methods using relatively large ensembles of simulations ($\gtrsim 50$) and a diverse set of initial conditions. Where the results disagree it may be ambiguous which approach is more accurate, therefore I focus on pure powerlaw models which should evolve self-similarly. This allows powerful self-consistency checks of the simulation results, since each output should, in a statistical sense, resemble scaled versions of earlier and later outputs. Fig. 1 shows an illustration of this, comparing snapshots from cosmological N-body simulations (initially $n = -1$ power spectrum) at three different times in the left panels and on the right showing a zoomed-in view of the previous output. Each image is a slice through 1/8th of the simulation volume and in each frame particles are randomly sampled from the simulation so that each side-by-side panel compares the same number of particles. Importantly, the non-linear scale, R_{nl}, which grows with time, is shown on each plot. The zoomed-in plots on the right have the same value of R_{nl}, relative to the axes shown, as the simulation outputs immediately to the left[2]. Qualitatively, for example in the sizes of voids and large structures, the zoomed-in slices on the right bear a remarkable resemblance to the simulation outputs shown adjacently to the left[3]. More rigorously, one can quantify the differences between the two images in a statistical way and in the limit of many realizations the scaled-forward density field should be statistically equivalent to the density field at the later output. To the extent that this happens the simulation can be said to evolve with self-similarity and it is this property of power-law initial conditions that I exploit throughout to evaluate the accuracy of both the conventional and ξ-sampled methods.

These kinds of "self-similar" tests were decisive in confirming the accuracy of the first generation of fully cosmological N-body codes [22]. I also show a few tests where, instead of a pure powerlaw, I simulate initial conditions consistent with a configuration space powerlaw times a gaussian bump. Investigated in [23] in great depth as a simplified model of BAO, this test is self-similar in a different sense – namely that the evolution of the dark matter clustering should only depend on the ratio of the scale of non-linearity to the scale of the BAO. I include these initial conditions as another test of the ξ-sampled method and as a valuable cross-check for the conventional method's predictions for the shift of the BAO feature.

I test these models extensively, focusing on pure powerlaw models with spectral slopes of $n = -1$, -1.5, and -2, and on the three models explored in [23] which resemble $n = -0.5$, -1, and -1.5 powerlaws in fourier space. I compare predictions from the two methods, showing results for the matter correlation function in § III, the power spectrum in § IV, and statistics of the halo population in § V. In § VIII I discuss my results and in § IX I summarize my main conclusions. Two appendicies discuss the effects of the integral constraint measurement bias on the correlation function and halo clustering bias, and another appendix compares results presented in § IV

[2] In Fig. 1, R_{nl} is defined as the scale where the rms overdensity in randomly placed spheres of radius, R, reaches the value of one. Different definitions for the non-linear scale are used elsewhere in the paper. This convention was chosen merely for visual appearance since, although each non-linear scale grows with the same rate, at a given epoch R_{nl} is a larger value than other choices.

[3] If unconvinced of this, look at the zoomed-in density fields on the right and try to tell if this density distribution is a flipped or mirrored version of the plot immediately to the left. If it is hard to tell on first glance this is evidence that the two images are statistically similar. Note that, formally, this is not possible with ΛCDM initial conditions but, largely because the ΛCDM power spectrum is roughly a powerlaw in simulations using boxes smaller than a few hundred Mpc, the eye could fairly easily be tricked if an analogous plot were made. Fig 1 is only intended as a pictorial illustration of the self-similarity of a cosmological density field.

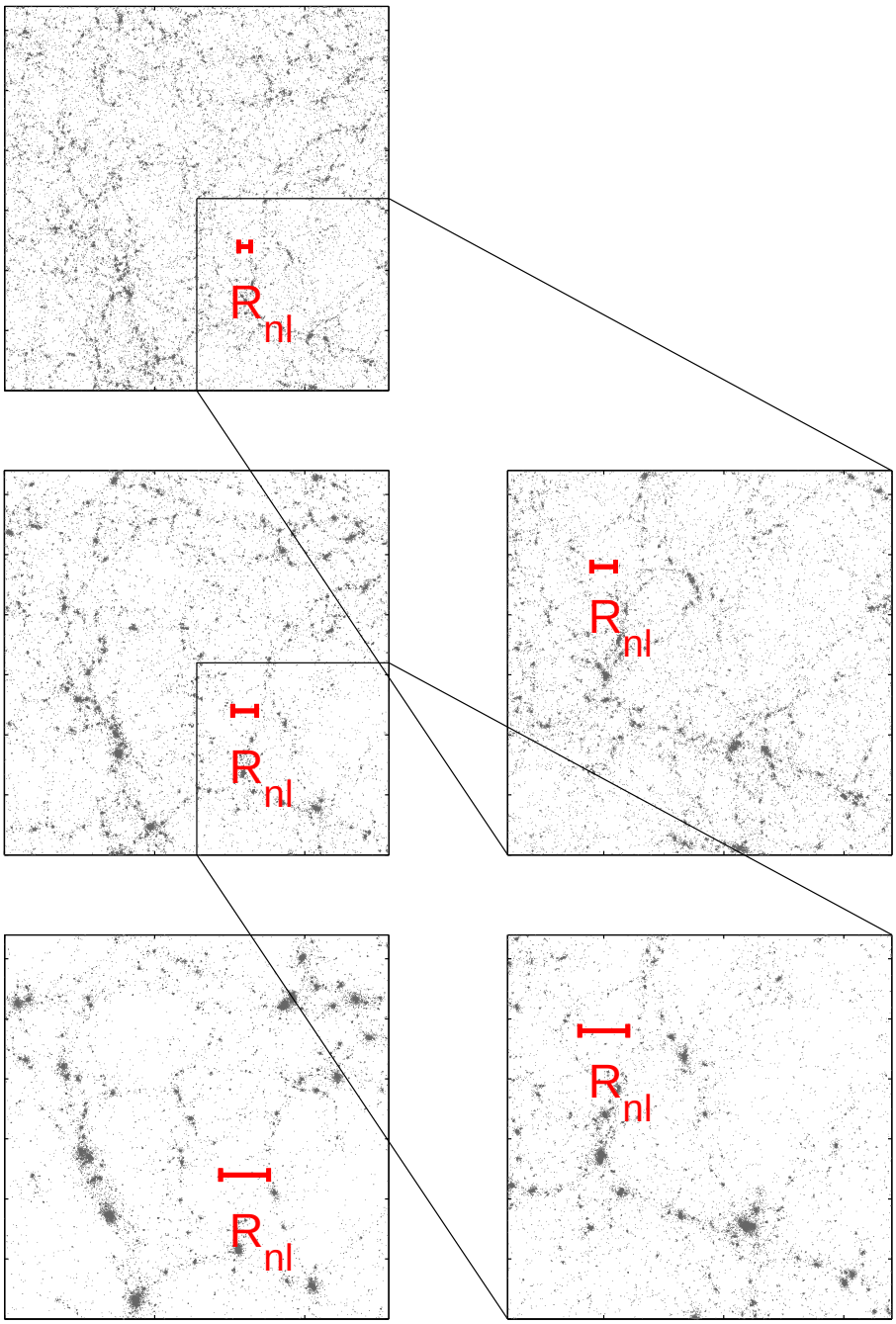

FIG. 1. *Left panels* Slices from an N-body simulation showing particles in the lower 1/8th of the simulation volume. The axes extend out to the box scale and the plots appear in chronological order – the earliest output at the top and the last output at the bottom. The nonlinear scale, $R_{\rm nl}$, is shown in each plot. *Right panels* Slices from the lower-right quadrant of the preceding output. The non-linear scale is shown on each plot, indicating that the slices have the same level of clustering as the outputs shown to the left.

to a perturbation theory model for the suppression of the growth of structure caused by the finite size of the simulation box.

II. INITIAL CONDITIONS

A. Overview of the ξ-sampled Method

In the ξ-sampled method implemented by [8], the (real space) matter correlation function for a given cosmological model is the fourier transform of the power spectrum

$$\xi(r) = \int \frac{d^3k}{(2\pi)^3} P(k)\, e^{i\vec{k}\cdot\vec{r}} = \frac{1}{2\pi^2}\int_0^\infty P(k)\,\frac{\sin kr}{kr}\,k^2\,dk. \quad (1)$$

To convolve $P(k)$ such that the simulated $\xi(r)$ is an exact match to Eq. 1 for $r < L_{\rm box}/2$, but is zero for larger separations, one simply fourier transforms $\xi(r)$ while cutting off the integral at $L_{\rm box}/2$ since $\xi(r) = 0$ for $r > L_{\rm box}/2$,

$$P_{\rm real}(k) = 4\pi\int_0^{L_{\rm box}/2} \xi(r)\,\frac{\sin kr}{kr}\,r^2\,dr. \quad (2)$$

I will refer to this result as $P_{\rm real}(k)$ to emphasize that this power spectrum is designed to maintain correspondence with the real space properties of the cosmological density field. Importantly, $P_{\rm real}(0)$ can be non-zero even if $P(0) = 0$; this term sets the fluctuations in the DC mode. In Appendix A of [8], using the subscript "uni" to denote variables in the model of interest and "box" to identify the parameters of the simulated volume, these fluctuations are mapped self-consistently onto fluctuations in cosmological parameters,

$$H_{0,\rm box} = H_{0,\rm uni}\frac{1}{1+\phi}, \quad (3)$$

$$\Omega_{m,\rm box} = \Omega_{m,\rm uni}(1+\phi)^2, \quad (4)$$

$$\Omega_{\Lambda,\rm box} = \Omega_{\Lambda,\rm uni}(1+\phi)^2, \quad (5)$$

$$\phi = \frac{5}{6}\frac{\Omega_m}{D(1)}\Delta_0, \quad (6)$$

where Δ_0 is a gaussian variable with mean zero and variance $P_{\rm real}(0)/L^3_{\rm box}$ and $D(1)$ is the value of the linear growth function at the present epoch. Note that Eq. 3 implies that in h^{-1} length units the box size of each simulation varies with the value of ϕ, whereas in standard length units (e.g. Mpc) the box size remains fixed. Similarly the box integrated mass, $M_{\rm box} = \rho_m L^3_{\rm box}$, varies from box-to-box in $h^{-1}M_\odot$ units, but is fixed in $M_\odot$ units.

Of crucial importance in deriving Eqs. 3-6 is the relationship between the scale factor of interest, $a_{\rm uni}$, and the corresponding scale factor in a particular realization, $a_{\rm box}$. In [8] this relationship is set by an approximate formula which determines $a_{\rm box}$ as the epoch where the age of the universe in the box is the same as the age of the unperturbed universe during the epoch of interest,[4]

$$a_{\rm box} \approx a_{\rm uni}\left(1 - \frac{1}{3}\frac{D(a)}{D(1)}\Delta_0\right). \quad (7)$$

[8] justified this formula by arguing that the ratio of the average density of the universe to the average density of a given box, $\bar{\rho}_{\rm uni}/\bar{\rho}_{\rm box} = a^3_{\rm box}/a^3_{\rm uni}$, is simply related to the overdensity of the box, which grows according to the linear theory growth function. Eq. 7 can also be obtained by Taylor expanding the perturbed $H(a_{\rm box})$ for small ϕ and equating the age of the universe in the box to the age of the universe at the epoch of interest.

B. Integration of Particle Trajectories

Having set up the initial conditions, determined the perturbed cosmological parameters of a given realization and computed the relevant scale factors, $a_{\rm box}$, for the epochs of interest, the initial conditions can be evolved using any cosmological N-body code. I use the publicly-available Gadget2 code with no modifications [24]. As a hybrid Tree-based code with PM grid for large scale forces, Gadget2 is a highly scalable N-body code which compares well to other codes used in the literature [e.g. 4]. Unless otherwise noted I show results from simulations with 256^3 particles and a 512^3 PM grid. Initial redshifts were set using $\Delta^2(k_{\rm Ny}) \lesssim 0.001$ as a rule of thumb [25], and the force softening was set to 1/20th the initial mean interparticle spacing.

C. Statistics of the Cosmological Density Field

With ensembles of simulations in the conventional method, the measurements of dark matter clustering at a given output, $a_{\rm uni}$, can typically be combined, and the statistical precision improved, with a simple average. In ξ-sampled simulations this procedure involves the extra step of weighting the measurements in each realization by factors of $a^3_{\rm box}$. Measuring $\xi(r)$ (in any context) is equivalent to calculating [26]

$$\xi(r) = \frac{\Gamma(r)}{\bar{n}} - 1 \quad (8)$$

where $\bar{n}$ is the average number density of particles in the volume and $\Gamma(r)$ is the number density of particles separated by length r. For an output from given realization using ξ-sampled ICs I am interested in $\xi_{\rm uni}(r) =$

[4] [11] was the first to appreciate that $a_{\rm box} \neq a_{\rm uni}$ but instead proposed to set $a_{\rm box}$ by matching the amplitude of the linear growth function in the perturbed cosmology. I return to this issue in VIII

$\Gamma(r)/\bar{n}_{\rm uni}-1$, and since $\bar{n}_{\rm uni}=\bar{n}_{\rm box}(a_{\rm box}/a_{\rm uni})^3$ this scaling is simply

$$\xi_{\rm uni}(r)=\left(\frac{a_{\rm uni}}{a_{\rm box}}\right)^3\frac{\Gamma_{\rm box}(r)}{\bar{n}_{\rm box}}-1=\left(\frac{a_{\rm uni}}{a_{\rm box}}\right)^3(\xi_{\rm box}(r)+1)-1 \tag{9}$$

Note that in Eq. 9 I have omitted an extra factor of $-2(a_{\rm uni}/a_{\rm box})^3$ that appears in [8]. I am unable to reproduce the results in [8] including this term, which could arise from using optimal estimators like [27] or [28]. I use the "natural" estimator, $\xi(r)=\Sigma_i DD_i/RR_i-1$, with RR_i derived analytically in the conventional way by multiplying the mean number density of particle pairs by the volume of the spherical shell. [8] does not state whether this or some more sophisticated estimator is used.

Another important subtlety to correlation function estimation worth mentioning is that for P-sampled simulations, an integral constraint is is artificially imposed on the scale of the box (e.g. [23, Appendix B]). It was noticed in re-producing Fig. 9 of [8], that the P-sampled correlation function results presented there were computed without proper correction for this measurement bias. As discussed in Appendix A, this negates part of the claimed advantage of the ξ-sampled method, which had agreed much better with the expected linear theory result on scales approaching the box.

Note that the integral constraint – which arises from assuming incorrectly that the average number density of pairs in the volume $L^3_{\rm box}$ is equal to the cosmic value – does not apply to ξ-sampled simulations because the mean density varies self-consistently from box to box. I will return to this issue in § VI, pointing out the close resemblence of the integral-constraint correction (Eq. A3) to the $(a_{\rm box}/a_{\rm uni})^3$ factor in Eq. 9.

The box-to-uni correction factor for the power spectrum (in ξ-sampled simulations) is determined straightforwardly from the fourier transform of Eq. 9,

$$P_{\rm uni}(k)=\left(\frac{a_{\rm uni}}{a_{\rm box}}\right)^3\langle\delta(\vec{k})\delta^*(\vec{k})\rangle\ . \tag{10}$$

If computed using h Mpc^{-1} units in k-space, the measurement of the bracketed term in Eq. 10, and the book keeping required to count of the number of modes contributing to each k for error estimation can be a fairly complicated exercise since the size of the simulation box varies in h^{-1} length units. In practice, it is much simpler (bearing in mind that box size is fixed in Mpc units) to convert the particle positions to Mpc units before applying the standard methods of mapping the particles onto a gridded density field [29, I use CIC] and computing its fourier transform. Similarly, measuring the halo mass function is considerably simplified by converting the particle positions to Mpc units and the particle masses to $M_\odot$ units (from $h^{-1}M_\odot$ units) before using standard halo finding methods. I use the FOF halo finder [30, 31] with a linking length of 0.2 times the mean interparticle spacing.

D. ξ-sampled ICs with Powerlaw Models

For powerlaw models, where $P(k)=Aa^2k^n$, the task of computing Eq. 2 is made substantially easier because an exact analytic solution for $\xi(r)$ is known in this case,

$$\xi(r)=\left(\frac{r_0}{r}\right)^{n+3},\ A\,a^2=\frac{2\pi^2\,r_0^{n+3}\,(2+n)}{\Gamma(3+n)\sin((2+n)\pi/2)}\ , \tag{11}$$

[32]. Eq. 2 therefore becomes

$$P_{\rm real}(k)=4\pi r_0^{n+3}\int_0^{L_{\rm box}/2}r^{-(n+1)}\,\frac{\sin kr}{kr}\,dr \tag{12}$$

and the DC power can be straightforwardly expressed as

$$P_{\rm real}(0)=4\pi r_0^{n+3}\int_0^{L_{\rm box}/2}r^{-(n+1)}\,dr \tag{13}$$

$$=\frac{2^{n+2}\pi}{-n}\left(\frac{r_0}{L_{\rm box}}\right)^{n+3}L^3_{\rm box}. \tag{14}$$

Analytic and special-function solutions to Eq. 12 exist for certain powerlaws. In this study I am interested in $n=-1$, -1.5 and -2 which can be expressed by

$$P_{\rm real,n=-1}(k)=\quad 4\pi r_0^2\,{\rm Si}({\rm kL_{box}}/2)\,{\rm k}^{-1}, \tag{15}$$

$$P_{\rm real,n=-1.5}(k)=2^{5/2}\pi^{3/2}r_0^{3/2}\,{\rm S}(\sqrt{\rm kL_{box}}/\sqrt{\pi}){\rm k}^{-1.5} \tag{16}$$

$$P_{\rm real,n=-2}(k)=\quad 8\pi r_0\,\sin^2(kL_{\rm box}/4)\,k^{-2}, \tag{17}$$

where Si(x) is the sine integral and S(x) is a Fresnel integral. These formulae can be very useful for generating accurate initial conditions, especially for steep power spectra. I show these power spectra in Fig. 2, fixing $r_0/L_{\rm box}=1/16$ to set the relative amplitudes. Notice that steeper powerlaws have larger DC power, easily seen on the plot as the asymptotic value of $P_{\rm real}(k)/L^3_{\rm box}$ for small k. Noticing that $P(k)$ does not go to zero at small k for the P-sampled powerlaws, one might be concerned that these models are unphysical. However, despite the high levels of large scale clustering power the rms overdensity in spheres and other statistics remain finite for $n>-3$.

E. Scale free?

Although pure powerlaw models are often referred to in the literature as "scale free," since $P(k)=Ak^n$ is featureless, the ξ-sampled initial power spectra shown in Fig. 2 clearly depend on the choice of $L_{\rm box}$. In practice, these oscillatory features die away in simulations and the effect of $L_{\rm box}$ is merely to change the variance of the DC mode (which is set by $P_{\rm real}(0)/L^3_{\rm box}$).

Since dark energy does introduce a new scale into the problem (e.g. the age of the universe when $\rho_m=\rho_\Lambda$), I consider only $\Omega_{m,{\rm uni}}=1.0, \Omega_{\Lambda,{\rm uni}}=0, \Omega_{k,{\rm uni}}=0$ so as to keep the simulations as "scale free" as possible and allow the self-similar tests discussed in the next section. In

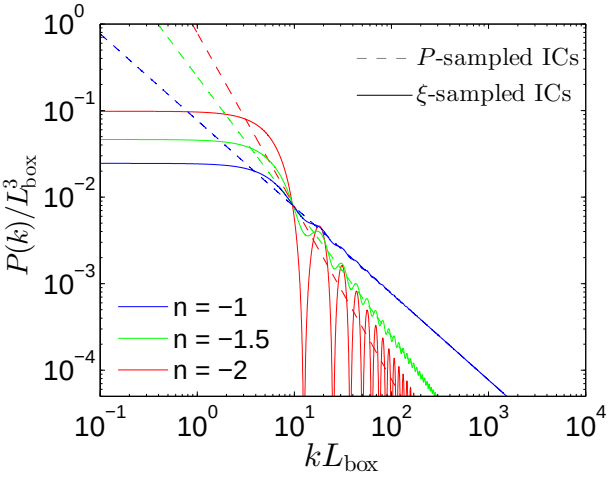

FIG. 2. A comparison of P-sampled and ξ-sampled pure powerlaw models. ξ-sampled power spectra are computed from Eq. 2 and used to generate initial conditions for that method. $r_0/L_{\rm box} = 1/16$ is chosen to set the overall amplitude of each model.

the Zeldovich [14] and adhesion [33, 34] approximations (as in linear theory), the effect of dark energy on structure formation is entirely captured by changing the linear theory growth function. [35] and [36] convincingly argue that this approximation is remarkably accurate even in the non-linear regime – the second order effect of dark energy is relatively small (see also Chapter 2). Therefore the results of my $\Omega_m = 1$ tests should still be quite relevant to studies that include a dark energy component.

As one final comment on the scale-free nature of my simulations, throughout I adopt, as a time variable,

$$\frac{a}{a_*} = \left(\frac{k_{\rm box}}{k_{\rm NL}}\right)^{(n+3)/2}, \qquad (18)$$

where $k_{\rm box} = 2\pi/L_{\rm box}$ and $k_{\rm NL}$ is defined by the dimensionless linear theory power spectrum, $\Delta^2(k_{\rm NL}) \equiv 1$. This choice is simply related to the $\sigma_{\rm miss}$ formula of [37], which quantifies the missing power on the scale of the box in P-sampled simulations. I adopt Eq. 18 for ease of comparison with [38] and because the $\sigma_{\rm miss}$ formula in [37] would be inappropriately applied to ξ-sampled simulations, which have a turnoff in $P_L(k)$ near the box scale (Fig. 2).

III. $\xi(r)$ RESULTS

A. Powerlaw Models

Fig. 3 shows my primary results for the self-similar scaling of the matter correlation function. The x-axis is shown in r/r_0 units where $\xi_L(r_0) \equiv 1$. Insofar as the dark matter clustering is negligibly affected by numerical limitations such as the finite scale of the box or the scale of the force softening, with this scaling the correlation function results from different outputs should all lie upon the same line. To the extent that this is achieved the correlation function can be said to evolve with self-similarity and it is clear from Fig. 3, excluding the first outputs which are severely affected by transients from initial conditions, that over a wide range of scales the results from these relatively modest, $N = 256^3$, simulations do fall upon the the same locus as expected. This locus is different for each powerlaw; for steeper power spectra (e.g. $n = -2$) power is transferred from large scales to small scales and the non-linear growth of $\xi(r)$ out paces linear theory whereas for shallower power spectra (e.g. $n \gtrsim -1$) there is so much small scale power that the process of halo formation and collapse causes the non-linear growth to fall behind linear theory in a process sometimes called "pre-virialization" [28]. In the language of the halo model [39] this implies that the predicted linear theory clustering on small scales is so high that the amplitude of the 1-halo term is below the linear theory clustering amplitude on those scales. The $n = -1$ case falls between these two extremes and the amplitude of the correlation function is both above and below linear theory, depending on the regime. (For a bracketing case of an even shallower power spectrum see, e.g., the $n = -0.5$ results in [23, Appendix A].)

In Fig. 3, the ξ-sampled and P-sampled methods generally agree on the shape of the self-similar solution. This is significant for the ξ-sampled results, on some level verifying the method. Alongside the measurements in each case the results from higher resolution simulations are shown. For $n = -1$ and $n = -2$ this comparison is made by numerically fourier transforming the non-linear power spectrum fitting functions published in [38]; note in the $n = -1$ case I include subtle but important corrections to their fit at small $k/k_{\rm nl}$ as determined in [23, Appendix A]. For $n = -1.5$, I compare with $\xi(r)$ measurements from 10 P-sampled simulations with $N = 512^3$ [23, Appendix A]. These high-resolution results are used more quantitatively in Fig. 4.

Important to point out in Fig. 3 is that the error on the mean $\xi(r)$ for the ξ-sampled method is much larger than the P-sampled simulations, despite running 2.5× as many ξ-sampled simulations in the $n = -1$ and $n = -2$ cases to improve the statistics, and 7.5× as many simulations in the $n = -1.5$ case to further ensure that the mean of the distribution of DC modes in the simulation set was very close to zero. For the ξ-sampled $n = -1$ case there are not enough realizations (50) to ensure that the mean of the distribution of DC modes agree with zero at the necessary level of accuracy (a criterion which will be discussed in § VII), and as a result, although difficult to see from the large error bars on the mean, there are strong deviations between the measurements of the mean $\xi(r)$ and the self-similar fit at large r for each output. The effect is more clearly illustrated in Fig. 4, which shows the agreement between the $\xi(r)$ measurements from my simulation set and the high-resolution self-similar solutions

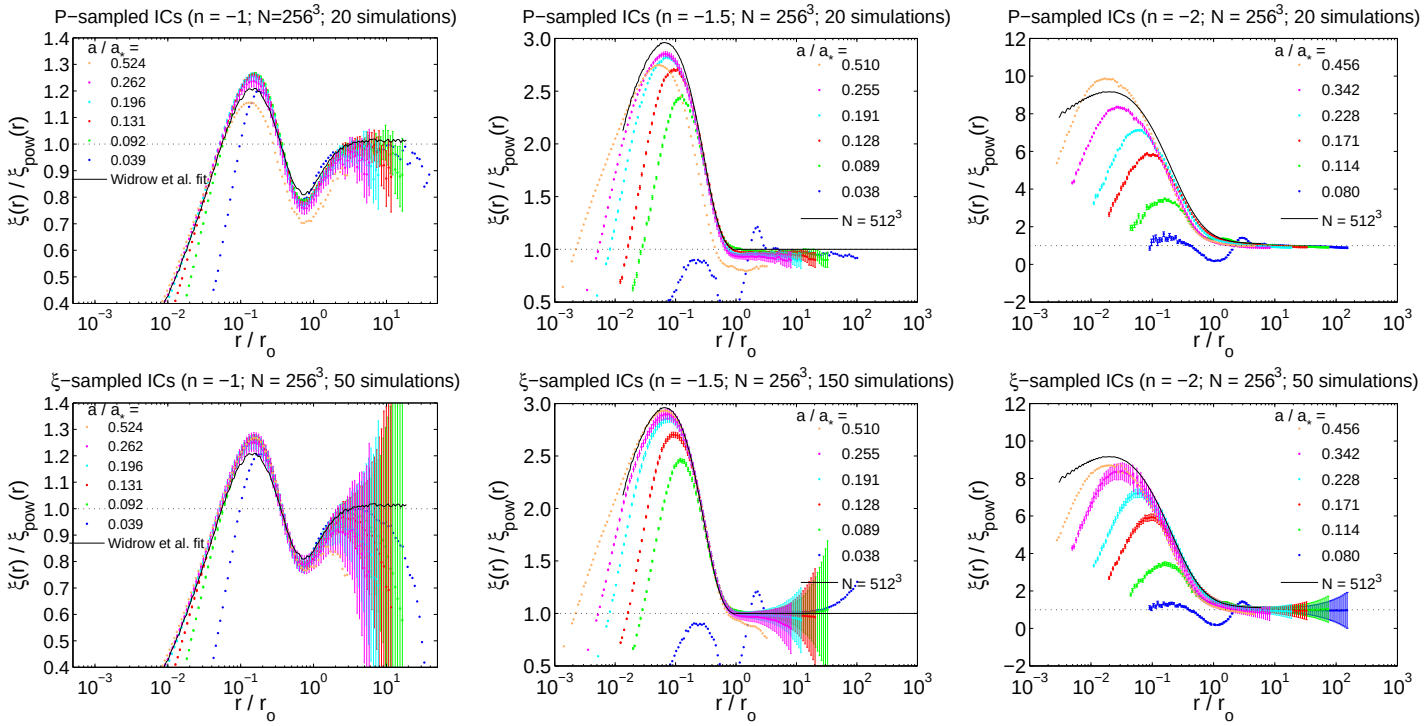

FIG. 3. Measured matter autocorrelation functions from conventional P-sampled (upper panels) and ξ-sampled (lower panels) ensembles of simulations. The left-most panels show results from an initially $n = -1$ power spectrum, middle panels show results from $n = -1.5$, and the right hand panels show $n = -2$. In each plot the x-axis is scaled by the non-linear scale, r_0, where $\xi_L(r_0) \equiv 1$ so that, if evolving with the expected self-similar behavior, the outputs should lie upon the same locus of points. The y-axis is scaled by $\xi_L(r) = (r_0/r)^{n+3}$. Black lines show high resolution results from P-sampled simulations for comparison (see text). Error bars show measured error on the mean. *Note that the first outputs in each plot are severely affected by transients from initial conditions.*

more quantitatively. Near $r/L_{\rm box} \sim 10^{-1}$ and independent of output, the ξ-sampled $n = -1$ results systematically fall well below the self-similar solution.

With this well-understood caveat, the agreement with the high-resolution self-similar results is fairly good and *excluding the initial and final outputs in each case* my simulation set tends to match the self-similar evolution to better than about 5% in most outputs and on most scales. The last output is excluded from this conclusion since the linear theory clustering level is so high that one expects departures from the true non-linear clustering from the suppression of power on the scale of the box. Also, the correction for the integral constraint, which assumes a *linear theory* correlation function in $\bar{\xi}(R_{\rm box})$, likely becomes inaccurate in this regime as well (Appendix A).

On scales much smaller than the box size, Fig. 4 shows that as structure evolves in the simulations the self-similar behavior extends below the scale of the initial mean inter-particle spacing, in some cases approaching the force softening. Since at best the initial conditions of the simulations only match the self-similar correlation function down to the initial mean interparticle spacing this result is non-trivial and difficult to anticipate from first principles. [40], using $n = -1$ simulations, show that Fourier modes in the non-linear regime are largely determined by the collapse of large-scale modes rather than by evolution of power initially on those scales. However, [41] and collaborators have argued that the common practice of setting the force softening significantly smaller than the initial mean interparticle spacing introduces errors which arise from the possibility that with this choice the equations of motion for the particles are no longer true to the Vlasov-Poisson fluid equations. Despite this, they also find that $\xi(r)$ (and $P(k)$) can reliably be modeled below the scale of the mean interparticle spacing. The main effect of aggressive force softening is to cause $\sim 5\%$ disagreement with the true non-linear $\xi(r)$ (and $P(k)$) on scales *larger* than the mean interparticle spacing. The P-sampled results shown in Fig. 4 are in qualitative agreement with the simulations presented in [41] in the sense that accurate non-linear behavior is observed below the mean interparticle spacing and on larger scales the measurements are consistent with the self-similar solution also at the level of $\sim 5\%$. Although beyond the scope of this paper, it would be interesting to run the P-sampled simulation set with less aggressive force softening (e.g. half the mean interparticle spacing) to test if the measured error on the mean $\xi(r)$ (and $P(k)$, see § IV) is detectably smaller, as predicted in [41]. At any rate, for all three powerlaws the self-similar behavior extends

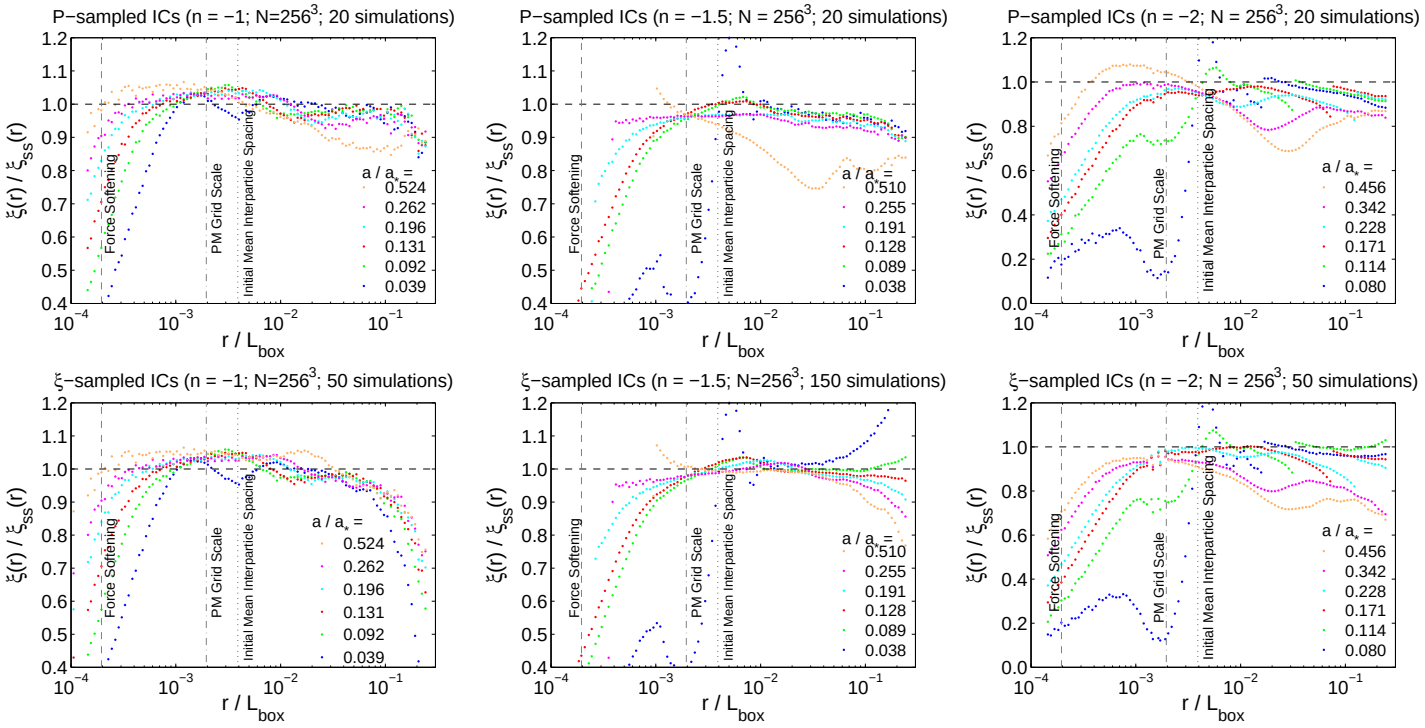

FIG. 4. Measured correlation functions from simulations relative to high-resolution results for the self-similar scaling ($\xi_{ss}(r)$; black lines in Fig. 3). Points show the mean $\xi(r)$. Panels are organized as in Fig. 3 (top panels: P-sampled results, bottom panels: ξ-sampled results, $n = -1, -1.5$ and -2 from left to right). Vertical lines show relevant *numerical* scales: the initial mean interparticle spacing (dotted black), the Particle-Mesh Grid (dot-dashed black), and the force softening (dashed black).

well below the scale of the mean interparticle spacing; it does not significantly depend on whether power is being rapidly "transferred" to smaller scales as for $n = -1.5$ and $n = -2$ or whether the non-linear growth proceeds less quickly than the linear theory prediction on small scales (i.e. $r < r_0$), as for $n = -1$.

B. Powerlaw Times a Bump Results

As discussed in depth in [23], a real-space powerlaw times a bump can be used as a self-similar numerical test in addition to providing insight into the non-linear physics of the evolution of the BAO feature. In this case,

$$\xi_{\rm IC}(r) = \left(\frac{r_0}{r}\right)^{n+3} (1 + A_{\rm bump}\, e^{-(r-r_{\rm bao})^2/2\sigma_{\rm bao}^2}), \quad (19)$$

and for resemblance to the ΛCDM correlation function I chose $A_{\rm bump} = 2.75$, $\sigma_{\rm bao}/r_{\rm bao} = 0.075$, and powerlaws of $n = -0.5$, -1, and -1.5. Unlike ΛCDM, this setup can be evolved much further than $\sigma_8 \sim 1$ to investigate the non-linear physics of the problem. I compare results in Fig. 5 from ξ-sampled ensembles of simulations with $N = 256^3$, $r_{\rm bao}/L_{\rm box} = 1/20$ to gaussian fits to the results of the P-sampled, $N = 512^3$, $r_{\rm bao}/L_{\rm bao} = 1/20$ sets of simulations in [23]. At least qualitatively the two simulation sets agree well. The slight small-scale discrepancy with the gaussian fit in the last output shown for the $n = -1$ case is a problem with the fit itself, as apparent in the P-sampled $n = -1$ results shown in Fig 5 while the discrepancies with the fit on scales larger than the bump stem from the mean of the fifty, randomly-sampled DC modes being slightly above zero, much like the pure powerlaw case for $n = -1$, a problem easily fixed by running more realizations. The slight mismatch in the amplitude of the bump in the $n = -1.5$ case is mirrored in the P-sampled, $N = 256^3$, $r_{\rm bao}/L_{\rm box} = 1/20$ results in Fig. 8 of [23], and the discrepancy, at large $r/r_{\rm bao}$, for the $n = -0.5$ case is in a regime where the amplitude of the correlation function is very weak and difficult to measure. Overall, the evolution of the bump is remarkably similar to previous results and, importantly, the non-linear shift of the bump in the $n = -1.5$ case and lack of a shift in the $n = -0.5$ and -1 cases agree with P-sampled simulations. Insofar as these two methods are equally valid approaches to setting up and running cosmological N-body simulations, this conclusion should be reassuring to the wider effort to characterize the non-linear shift of the BAO peak using standard, P-sampled simulations.

As with the powerlaw results in the previous section, despite running a few times more simulations than the P-sampled ensembles, the measured error on the mean $\xi(r)$ is still quite large. (In [23], I typically ran only seven realizations in each case.) This seems to be the case for a variety of powerlaws and in simulations with and without a bump. A close inspection of the original

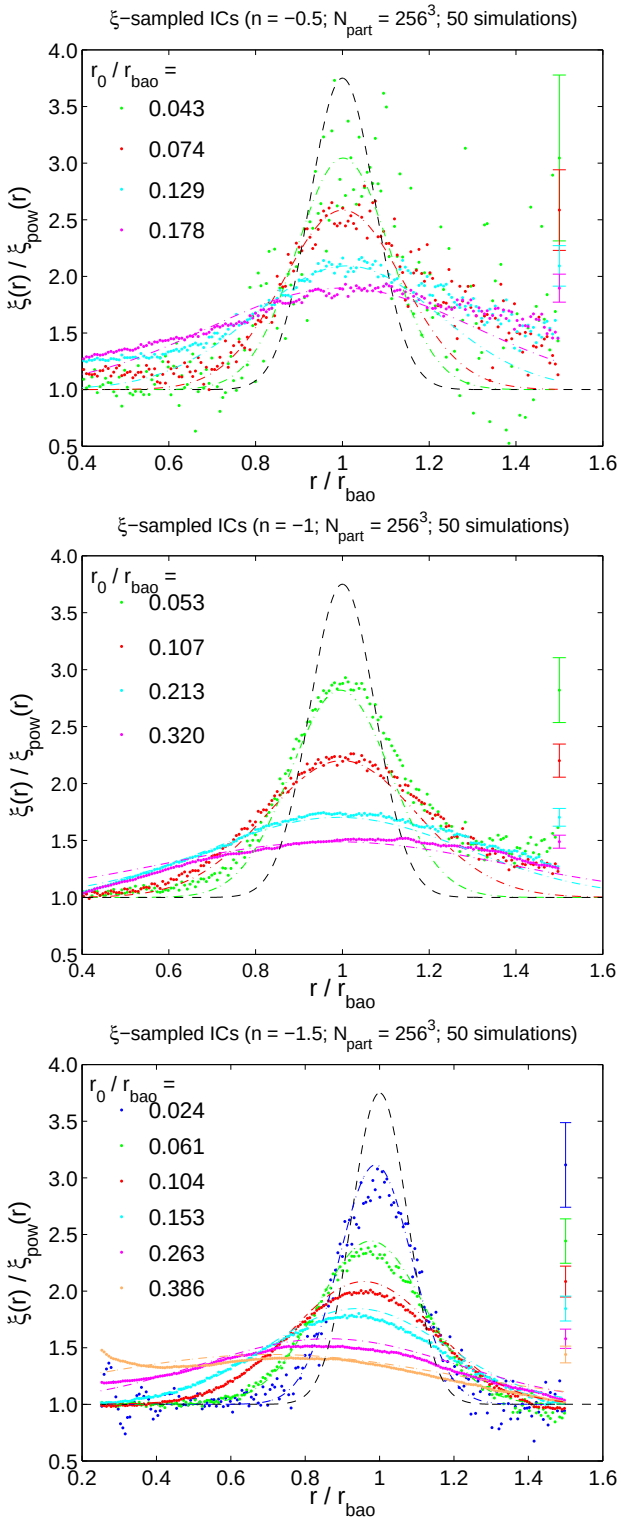

FIG. 5. Correlation function results from ensembles of 50 ξ-sampled simulations using initial conditions consistent with a powerlaw times a gaussian bump as a simplified model of baryon acoustic oscillations. Dot-dashed lines show gaussian fits to the high-resolution simulations presented in [23, Fig. 3]. Typical errors on the mean are shown offset to the right.

$\xi(r)$ results in [8] likewise show in ΛCDM simulations that the error on the mean $\xi(r)$ is significantly larger in ξ-sampled simulations. [8] does not not comment on this interesting result. I return to this issue in § VI.

IV. P(K) RESULTS

Although ξ-sampled simulations are designed to match the real-space properties of the cosmological density field, it is nevertheless interesting to measure the power spectra in these simulations as another self-similar numerical test. These results are presented in Fig. 6. As seen in the ΛCDM simulations in [8], the oscillations in the initial ξ-sampled power spectrum largely die away. These features are essentially absent in the $n = -1$ case while in the $n = -1.5$ and $n = -2$ cases the oscillations about linear theory for $k/k_{\rm NL} \ll 1$ seem to persist for longer. This is likely explained by the fact that there are more factors of expansion between the epoch of the initial conditions and an output at a given a/a_* in the $n = -1$ case than the $n = -1.5$ or $n = -2$ cases. (As a result, the $n = -1$ case requires much longer execution times to reach $a/a_* \sim 0.5$ than $n = -1.5$ or $n = -2$.)

I compare the power spectrum results in Fig. 6 to fitting functions from higher resolution simulations. For $n = -1$, I compare with an improved version [23, Appendix A] of the fitting function from the high-resolution simulations by [38]. The $n = -2$ fitting function comes directly from [38] and the $n = -1.5$ fitting function (which was not investigated in [38]) comes from 10 P-sampled, $N = 512^3$ simulations presented in [23, Appendix A]. Generally, the agreement with the higher resolution non-linear fits is quite good and, intriguingly, the final outputs of the ξ-sampled results in each case seem to agree more closely with the non-linear fitting function than the P-sampled case. One might have expected the ξ-sampled simulations to more accurately handle situations with large amounts of clustering power on the scale of the box, since the fluctuating DC mode is more true to the large scale clustering in the real universe. With the correlation function, it is difficult to see which scheme performs better in this regime, since a discrepancy between the P-sampled correlation function results from different outputs – a violation of self-similarity – or with some higher-resolution result can potentially be explained as a failure of the integral constraint correction, which assumes a linear theory model for $\xi(r)$ in $\bar{\xi}(R_{\rm box})$.

The self-similarity of the power spectrum exhibited in ξ-sampled simulations when the clustering levels become very high on the scale of the box appears to be one of the principal improvements offered by this approach in the tests shown here. Ironically, a method designed to match the real-space properties of the density field appears to more accurately render the fourier-space statistics in this regime. It is tempting to speculate that the difficulties, historically, with realizing self-similar evolution in $n = -2$ simulations with small numbers of parti-

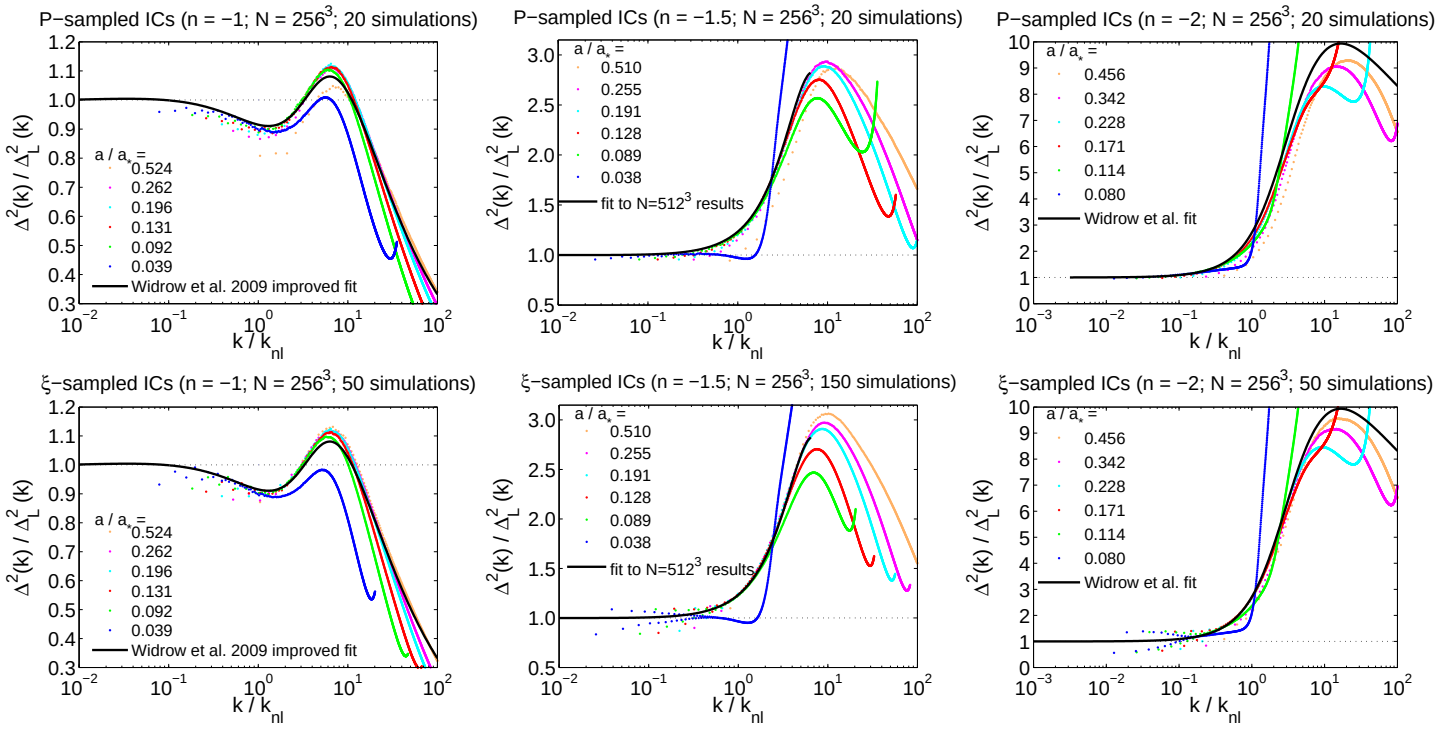

FIG. 6. Comparison of non-linear power spectrum results from P-sampled (top panels) and ξ-sampled (lower panels) ensembles of simulations for initially $n = -1$ (left), $n = -1.5$ (middle) and $n = -2$ (right) powerlaws. The x-axis is scaled by the non-linear wavenumber $k_{\rm nl}$, defined by $\Delta^2(k_{\rm nl}) \equiv 1$. Solid black lines show results from high-resolution simulations (see text).

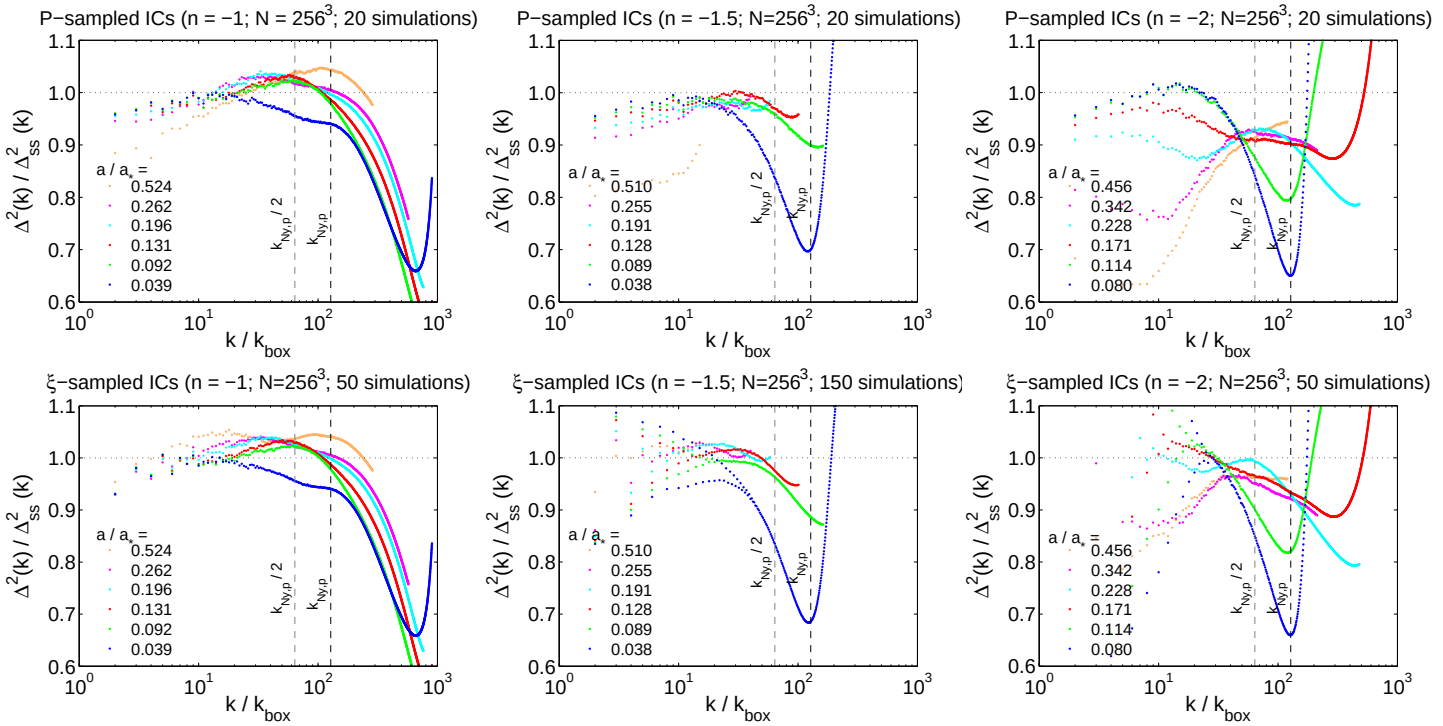

FIG. 7. Non-linear power spectrum results relative to fitting functions to results from high-resolution simulations ($\Delta^2_{ss}(k)$; black lines in Fig. 6). Panels are organized as in Fig. 6. The x-axis is shown relative to $k_{\rm box} = 2\pi/L_{\rm box}$ and dashed vertical lines show the particle nyquist wavenumber and half the particle nyquist wavenumber. The $n = -1.5$ results do not extend to large k because non-linear fitting functions from very high-resolution ($N \gg 512^3$) simulations are unavailable in this case.

cles could have been alleviated by this method [22, 42]. However, note that the errors on the mean in Fig. 6 are considerably larger in ξ-sampled simulations with even $2.5\times$ the number of realizations than in P-sampled simulations. This penalty is severe – at least doubling the computational expense to obtain the same errors on the mean power spectrum and correlation function. I return to this issue in § VI, discussing whether "paying" this penalty is more advantageous than increasing the numbers of particles in the box in simulating the non-linear evolution of very steep ($n \lesssim -2$) power spectra as would be relevant to studying the evolution of the very large k power spectrum in ΛCDM [43, 44].

As a more quantitative comparison in Fig. 7 I show the ratios of the non-linear power spectrum results to the high-resolution self-similar fits. Bearing in mind, again, that the initial power spectrum has oscillatory features from the cutoff in Eq. 2, the ξ-sampled simulations typically match the self-similar solution on scales close to the box better than the P-sampled simulations. On scales at or below the initial mean interparticle spacing ($k \gtrsim k_{\rm Ny,p}$) all the simulations show evidence that self-similarity extends into this regime. The result is most dramatic for the $n = -2$ results where the first output has very little spectral power at $k \approx k_{\rm Ny,p}$ but a number of later outputs agree with the self-similar scaling to better than 10%. A similar trend is apparent for $n = -1.5$ but the non-linear fit comes from much lower resolution simulations than $n = -2$ or -1 [23, Appendix A]. In Fig. 7 each output is only shown up to the range of validity of the high-resolution non-linear fitting functions. For $n = -1$, the first output shown was taken after the particles moved significantly from their initial positions, which is why there is appreciable spectral power already at $k \approx k_{\rm Ny,p}$. In all the simulation results the self-similar behavior extends continually further beyond the initial mean interparticle spacing as the simulation progresses. This echoes a result from [41], who, in focusing on ΛCDM simulations, state "We see this maximal resolved wavenumber propagates to larger k in time, reaching a final value of order [the force softening]". Unfortunately, the dynamic range of Fig. 7 does not extend all the way to the force softening scale at $k/k_{\rm box} \approx 2560$. This is primarily due to the limited range of the high-resolution non-linear fitting functions. To ameliorate the concern that the grid used to measure $\Delta^2(k)$ may be affecting the behavior at large $k/k_{\rm box}$, except for the ξ-sampled $n = -1.5$ results, a 1800^3 grid was used to measure $\Delta^2(k)$ from each realization. The measurement should therefore be reliable out to $k \approx k_{\rm Ny,grid}/2$ or $k/k_{\rm box} \approx 450$.[5]

[5] The ξ-sampled $n = -1.5$ results were computed with a 1024^3 grid to avoid the computational expense of performing an extra 150, 1800^3 FFTs in a case where the non-linear fitting function is the bottleneck at large $k/k_{\rm box}$. This is a minor detail and the P-sampled $n = -1.5$ results, which were computed with a 1800^3 grid, agree well with the ξ-sampled measurements at large $k/k_{\rm box}$.

Finally, to comment again on the small $k/k_{\rm box}$ behavior, the P-sampled results show clear trends of underpredicting the non-linear power spectrum as the clustering level approaches the box scale. Seto [45] proposed modeling the inaccuracies introduced from the finite scale of the box using perturbation theory but stopped short of testing this approach with cosmological N-body simulations. Since the investigation is largely orthogonal to the narrative of comparing the P-sampled and ξ-sampled methods to each other, this study, which finds qualitative agreement between the $n = -1.5$ and $n = -2$ results and the perturbation theory expectations, is presented in Appendix C.

V. HALO MASS FUNCTION AND HALO BIAS

As aspects of structure formation with their own set of numerical challenges [e.g. 25, 49], I also investigate the halo abundance and clustering bias in my simulation ensembles. I principally compare my results to [46] who measured the $f(\nu)$ mass function in a variety of different powerlaw models using higher resolution simulations than employed here but with far fewer realizations. Using halo abundances, [46] reports their best fit parameters for each powerlaw to the Sheth-Tormen [48] form of $f(\nu)$, a result which can be used to predict both the halo abundance and the halo clustering using the framework of [48]. Other models for $f(\nu)$ exist in the literature as calibrated fits to ΛCDM simulations. I principally compare to [46], who focused on pure powerlaw models.

A. Halo Mass Function

Fig. 8 shows my primary results for the halo mass function. To highlight the self-similar evolution in the upper panel of each sub-figure I multiply the halo mass function by a factor of $(4\pi/3)R_*^3 = M_*/\rho_m$ where $\sigma(M_*) = \delta_c$. This choice is motivated as follows: the halo mass function can be characterized solely as a function of $\sigma(M)$,

$$\frac{dn}{d\log M} = \frac{\rho_m}{M}\frac{d\log\sigma^{-1}}{d\log M}\nu f(\nu) \tag{20}$$

where $\nu = \delta_c/\sigma(M)$ and $f(\nu)$ is calibrated to simulations or estimated analytically. For powerlaw models $\nu = (M/M_*)^{(n+3)/6}$ and Eq. 20 becomes

$$\frac{dn}{d\log M} = \frac{\rho_m}{M}\left(\frac{n+3}{6}\right)\left(\frac{M}{M_*}\right)^{(n+3)/6} f\left(\left(\frac{M}{M_*}\right)^{(n+3)/6}\right), \tag{21}$$

so that in multiplying by $M_*/\rho_m = (4\pi/3)R_*^3$ this expression becomes a unique function of M/M_*. Alternatively, this $(4\pi/3)R_*^3$ factor can be thought of as a characteristic non-linear volume, the abundance of halos at a particular M/M_* halo mass being set by the ratio of the fixed volume to the characteristic non-linear volume. For this

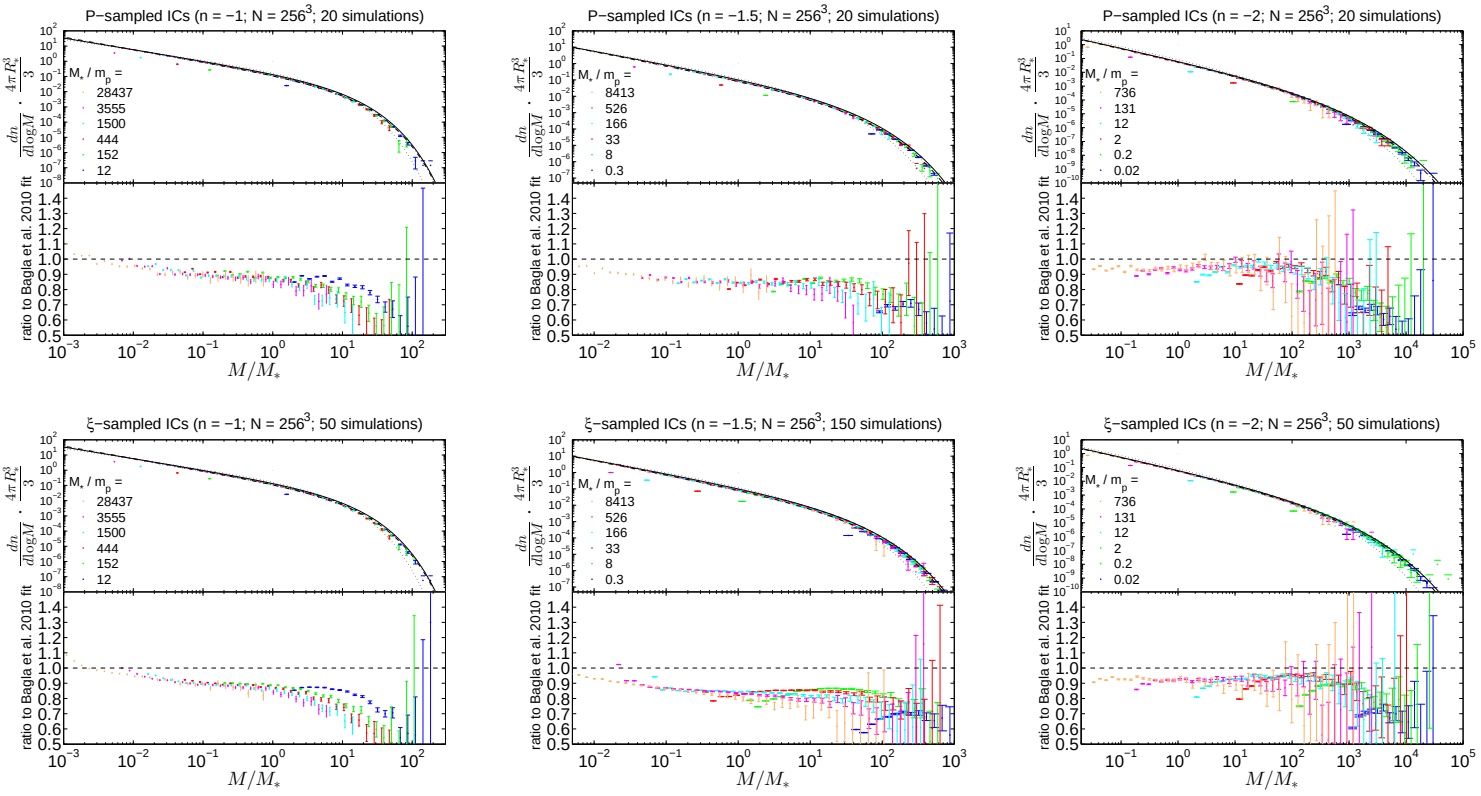

FIG. 8. Halo mass function results for my fiducial simulation set. Results are shown either scaled by $(4\pi/3)R_*^3 = M_*/\rho_m$ (upper sub-panels in each case) or relative to the fitting function of [46] (lower sub-panels in each case). For each output the characteristic non-linear mass ($\sigma(M_*) \equiv \delta_c$) relative to the mass of a particle in the simulation is shown in the legend; each output is colored consistently with other plots. Error bars show the error on the mean.

reason the earliest simulation outputs probe the large M/M_* regime because the fixed volume of the simulation is so much larger than $(4\pi/3)R_*^3$ that rare supermassive objects can be more easily found than at later outputs when the non-linear volume is significantly larger.

In Fig. 8 both ξ-sampled and P-sampled simulation sets give remarkably similar results in each case. Dividing by the prediction from [46] as in the bottom panels in each sub-figure does not break self-similarity, and the overlap of the mass functions from each output is still a test of self-similar evolution that my simulation results typically achieve, although, interestingly, the self-similar solutions deviate significantly from the fits reported in [46]. This can be explained by noticing in the plots shown in [46] that the measurements of $f(\nu)$ from their simulations can fall below the best fit $f(\nu)$ by up to 50%. Had they used more than two free parameters in their fit to $f(\nu)$ (as they did in adopting the Sheth-Tormen framework), it seems likely that my results would agree more closely with theirs.

B. Halo Bias

In the Sheth-Tormen [48] framework the halo bias is given by

$$b(M) = 1 + \frac{a\nu - 1}{\delta_c} + \frac{2p/\delta_c}{1 + (a\nu)^p} \tag{22}$$

where a and p are free parameters. Using the best fit values for a and p quoted in [46] for each powerlaw, in Fig. 9 I compare their results to the measured halo bias in my simulations. Although the halo bias can be equally well measured in fourier space [49, 50] as in real space, to avoid the oscillatory features in $P(k)$ I measure the bias with $b^2(M) = \xi_{hh}(r)/\xi_{mm}(r)$ in both P-sampled and ξ-sampled simulation results, where $\xi_{hh}(r)$ is the halo-halo correlation function and $\xi_{mm}(r)$ is the matter-matter correlation function (elsewhere referred to as $\xi(r)$). One subtlety with this measurement deserves mention: in P-sampled simulations the $\xi_{hh}(r)$ measurements, like the $\xi_{mm}(r)$ measurements, suffer from the artificial imposition of the integral constraint on the scale of the box. In Appendix B, drawing upon the procedure for correcting $\xi_{mm}(r)$ described in [23, Appendix B], I demonstrate that this correction cancels (at least to first order) in

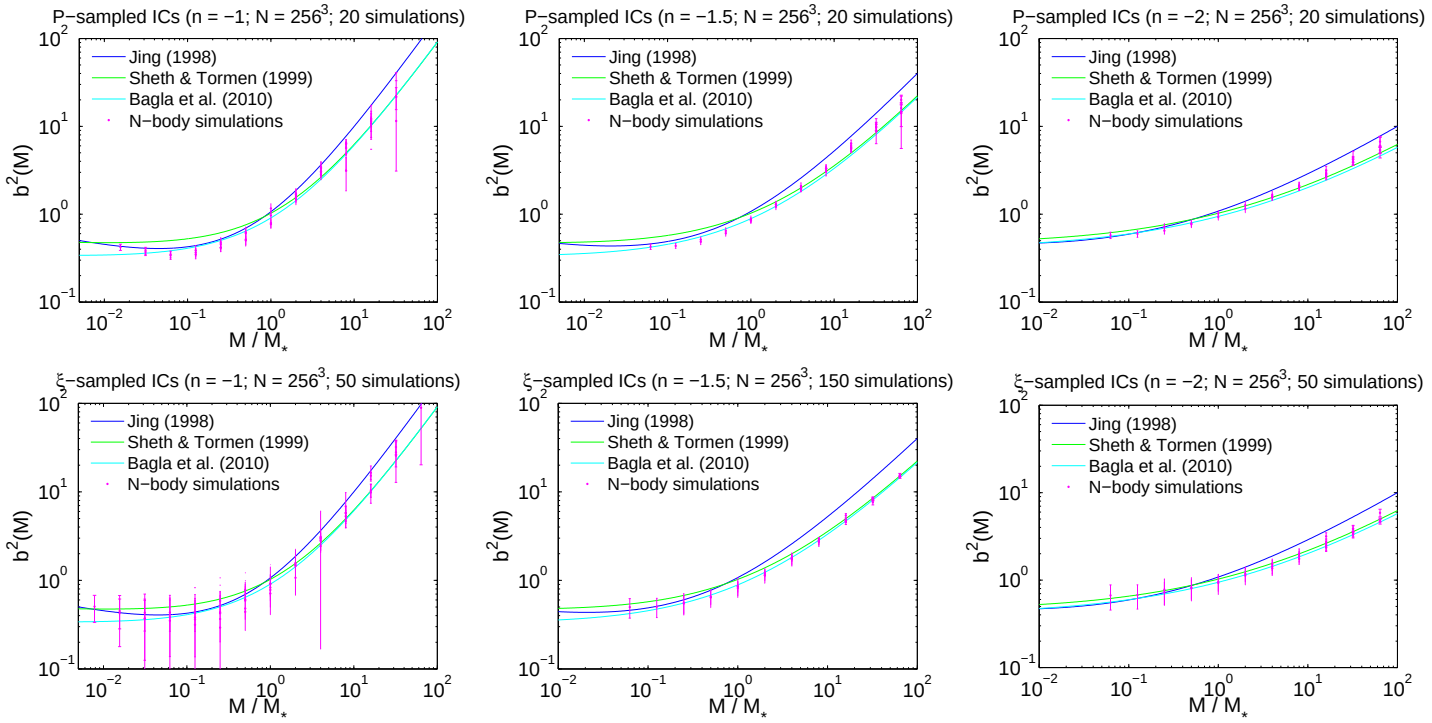

FIG. 9. Halo clustering bias measured with $b^2(M) = \xi_{hh}(r)/\xi_{mm}(r)$. Results from my simulations are shown by magenta points; error bars show the measured error on the mean. In each panel the bias functions of Jing [47] (blue), Sheth-Tormen [48] (green) and Bagla et al. [46](cyan) are shown for comparison. Since the Sheth-Tormen [48] bias has been calibrated to ΛCDM simulations (rather than scale-free models) this prediction is shown only for reference to the familiar ΛCDM case.

the measurement of the clustering bias. ξ-sampled simulations naturally circumvent this issue, and no integral-constraint corrections are needed to the measured $\xi_{hh}(r)$ and $\xi_{mm}(r)$.

Generally, in Fig. 9 the bias measurements from the two methods agree within errors and the results typically stay closer to the [46] predictions rather than the formula from [47]. (The Sheth-Tormen [48] bias, calibrated to ΛCDM simulations, is overplotted for comparison.) At large M/M_* in particular the measurements typically fall below the fitting function of [47], as previously observed for $n = -1$ in [36], who point out that the fit deviates slightly from the original simulations in [47]. My results (and indirectly the $f(\nu)$ results of [46]) corroborate this conclusion.

Of particular note in Fig. 9 the ξ-sampled bias for $n = -1$ shows substantially larger error bars at low M/M_* than the P-sampled case. This discrepancy stems from the large variance of $\xi_{mm}(r)$ and deviation from the self-similar fit, as seen in Figs. 3 & 4, rather than the $\xi_{hh}(r)$ measurement.

Taken together the results in this section assert that the process of halo formation is consistent between the two methods, at least for the measurements I show here (see also [19]).

VI. BOX-TO-BOX VARIANCE

Having explored the ensemble-averaged predictions for the mean $\xi(r)$, $P(k)$, and halo mass function, in this section I turn to comparing the box-to-box variance of these quantities with expectations from theory. As emphasized by [5], it is important to be able to model both the mean dark matter and/or halo statistics and the covariances of these quantities to infer the appropriate cosmological constraints from a set of large-scale structure data. In this vein, there are a number of groups that have taken up this concern for ΛCDM [6, 51–54]. By comparing gaussian and other expectations for the box-to-box variance of these statistics, I aim to compare these predictions to a more diverse range of cosmological models with the added simplicity and clarity of powerlaw setups.

A. Var ξ results

Fig. 10 shows my main results for the box-to-box variance of the correlation function, in each case multiplying by a well-motivated power of $(L_{\rm box}/r_0)$ to normalize the y-axis. In each plot I compare with the expected Gaussian variance for a volume $V = L_{\rm box}^3$ as well as an estimate of what I refer to as a "floor" to the variance from non-linear effects (i.e. higher-order statistics) pro-

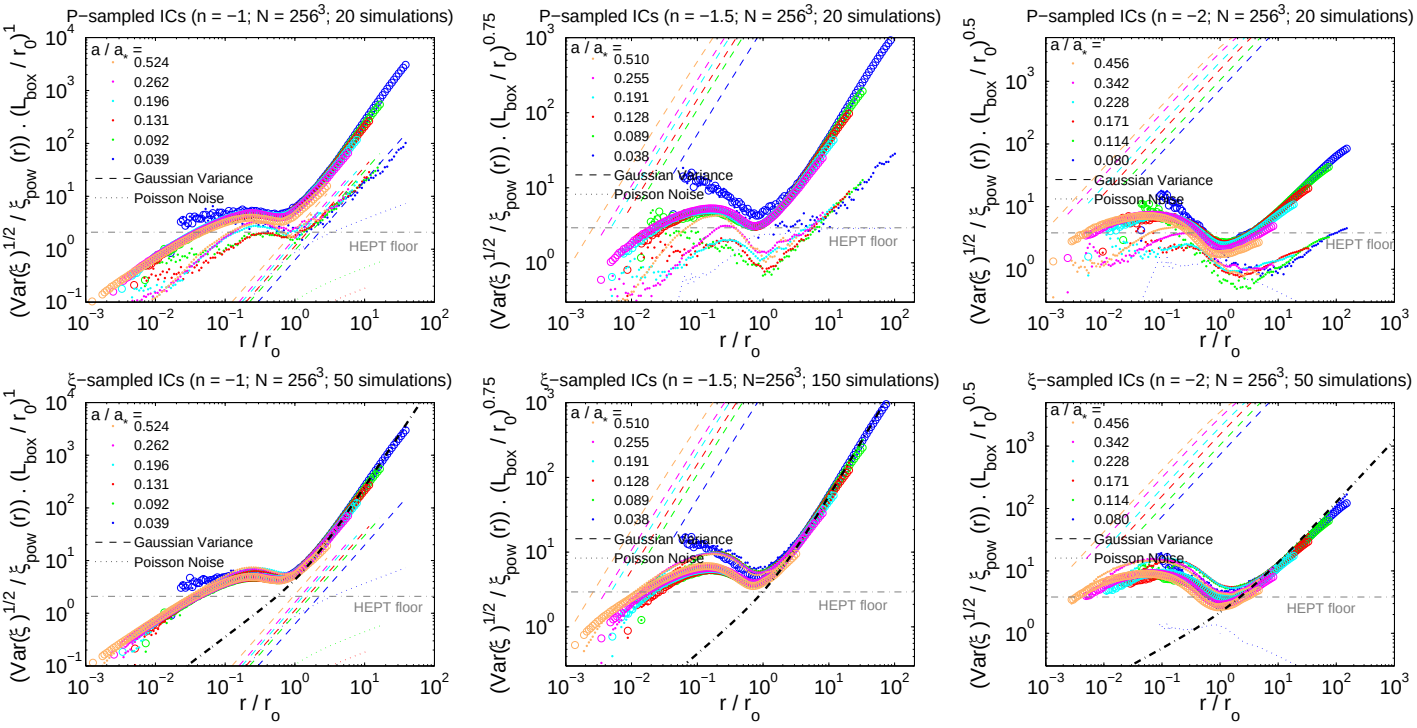

FIG. 10. Measurements of the box-to-box variance of $\xi(r)$ (colored points) compared to expectations from Gaussian statistics (dashed colored lines). Open circles show a jackknife estimate of the variance from breaking up each simulation box into octants and measuring the correlation function in each sub-volume. The anticipated contribution to the variance from higher-order correlations (as predicted by "Hyper-Extended"PT [55]) is shown with gray dot-dashed lines. The y-axis in each plot is scaled by $(L_{\rm box}/r_0)^{(n+3)/2}$ so that this expected "floor" to the box-to-box variance is a fixed horizontal line. Also shown is the Poisson noise level (colored dotted lines) which only happens to be large enough to be visible for the first outputs in each case and, for the ξ-sampled results in the lower three panels, a black dot-dashed line shows the result of Eq. 30.

vided by Hyper Extended Perturbation Theory [HEPT; 55]. This term produces a fractional error proportional to the volume averaged correlation function,

$$\frac{\sigma_{\xi,\rm hept}}{\xi_{\rm pow}(r)} = \sqrt{4(1-2Q_3+Q_4)\bar{\xi}(R_{\rm box})} \qquad (23)$$

where Q_3 and Q_4 come from HEPT[6]. Note that the scalings in Fig. 10 are designed to cancel with the $\sqrt{\bar{\xi}(R_{\rm box})} \sim (r_0/L_{\rm box})^{(n+3)/2}$ dependence of this term so that the "HEPT floor" is a horizontal line in each plot. Although not expected to be an extremely good match to simulations, I include it in Fig. 10 for qualitative comparison to my results.

The Gaussian variance is given by

$$\sigma_\xi^2 = \frac{1}{V\pi^2}\int_0^\infty dk k^2 \left(\frac{\sin kr}{kr}\right)^2 P(k)^2 \qquad (24)$$

[53], which, for powerlaw, $P(k) = Ak^n$, models can be shown to be equivalent to

$$\frac{\sigma_\xi}{\xi_{\rm pow}(r)} = \frac{A_n}{\pi}\frac{\sqrt{\Gamma(1+2n)\sin n\pi}}{4^{(n+1)/2}}\left(\frac{r}{L_{\rm box}}\right)^{3/2} \qquad (25)$$

where $A \equiv A_n r_0^{n+3}$, and $\Gamma(1+2n)$ is the usual gamma function. Notice that all of the r_0 dependence has canceled out with the division by $\xi_{\rm pow}(r) = (r_0/r)^{n+3}$. Unfortunately, Eq. 25 is only convergent for the limited range of $-1.5 < n < -0.5$. Therefore, the Gaussian variance predictions for the $n = -1.5$ and $n = -2$ results in Fig. 10 include cutoffs in $P(k)$ at $k = k_{\rm box}$. As a correction to a divergent result, the poor comparison of the expected Gaussian variance for these two cases is unsurprising.

In Fig. 10, the ξ-sampled results clearly exhibit much larger box-to-box variance than the P-sampled simulations. The ξ-sampled $n = -1$ results are perhaps most instructive since the Gaussian variance is accurate in this case – the measured variance is clearly much higher than this expectation. Though initially quite puzzling, the box-to-box variance of $\xi(r)$ turns out to be almost entirely dominated by the variance of the DC mode, so much so that it is as if the $(a_{\rm uni}/a_{\rm box})^3$ factor in Eq. 9 is the direct cause of the large variance over a wide regime

[6] The Journal version of this paper contains a typographical error in the equation for Q_4. The arXiv version is correct or, c.f., [56].

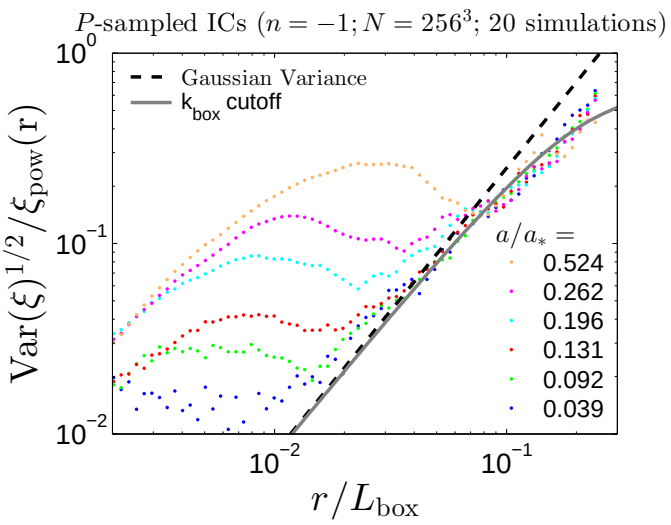

FIG. 11. Focusing on the box-to-box variance of $\xi(r)$ for the P-sampled, $n = -1$ simulation ensemble (top left plot in Fig. 10). The y-axis is shown without the $(L_{\rm box}/r_0)^{(n+3)/2}$ scaling used in Fig. 10 and the x-axis shows the scale, r, of the correlation function measurement relative to the box scale so that the expected Gaussian variance (dashed black line; Eq. 25) appears as a fixed line. Results from including a low k cutoff at $k = 2\pi/L_{\rm box}$ in the expected Gaussian variance are shown for comparison as a solid gray line.

without any interference from the variance of the $\Gamma(r)$ term. More explicitly, since $a_{\rm box,i}/a_{\rm uni} = 1 - \Delta_i/3$,

$$\left(\frac{a_{\rm uni}}{a_{\rm box,i}}\right)^3 \approx 1 + \Delta_i + \frac{2}{3}\Delta_i^2 \tag{26}$$

and referring to $\xi_{\rm uni,i}(r)$ as the result of scaling $\xi_{\rm box,i}(r)$ with the appropriate weighting,

$$\begin{aligned}\xi_{\rm uni,i}(r) &\approx (1 + \Delta_i + \frac{2}{3}\Delta_i^2)\frac{\Gamma(r)}{\bar{n}} - 1 \qquad (27)\\ &= \frac{\Gamma(r)}{\bar{n}} - 1 + \frac{\Gamma(r)}{\bar{n}}\Delta_i + \frac{2}{3}\Delta_i^2\frac{\Gamma(r)}{\bar{n}}\\ &= \xi_{\rm box,i}(r) + (1 + \xi_{\rm box,i}(r))\Delta_i + (1 + \xi_{\rm box,i}(r))\frac{2}{3}\Delta_i^2 \ .\end{aligned}$$

Ignoring, for the moment, the Δ_i^2 term,

$$\xi_{\rm uni,i}(r) \approx \xi_{\rm box,i}(r) + (1 + \xi_{\rm box,i}(r))\Delta_i \tag{28}$$

and computing the variance of $\xi_{\rm uni,i}$ while assuming that $\langle\Delta_i\rangle = 0$ (i.e. that the mean of the DC mode is equal to zero) it can be shown

$$\begin{aligned}\langle\xi_{\rm uni,i}^2\rangle - \langle\xi_{\rm uni,i}(r)\rangle^2 &\approx (1 + 2\langle\xi_{\rm box,i}(r)\rangle)\langle\Delta_i^2\rangle \qquad (29)\\ &= (1 + 2\langle\xi_{\rm box,i}(r)\rangle)\frac{2^{n+2}\pi}{-n}\left(\frac{r_0}{L_{\rm box}}\right)^{n+3} \qquad (30)\end{aligned}$$

where in the last step I used the fact that the variance of the DC mode, $\langle\Delta_i^2\rangle$, is equal to $P_{\rm real}(0)/L_{\rm box}^3$. Using black dot-dashed lines, I compare this formula, with much agreement, to the ξ-sampled results in Fig. 10. An important detail in this derivation, which was lost in dropping the Δ_i^2 term in Eq. 28, is that the *mean* correlation function (i.e. $\langle\xi_{\rm uni,i}(r)\rangle$) will still retain a $(2/3)\langle\Delta_i^2\rangle$ term. This term is schematically very similar to the integral constraint correction described in [23, Appendix B]. The difference is that in ξ-sampled simulations this correction is, in a sense, applied individually to each simulation with a large scatter, rather than statistically as a correction to the ensemble-averaged correlation function. The $n = -1$ case, where the expected variance of $\xi(r)$ from Gaussian statistics is still well defined, is decisive for assessing which approach, if either, is more accurate. With this in mind, the lower left panel of Fig. 10 shows that the variance from the ξ-sampled method substantially exceeds the Gaussian expectations and the upper left panel of Fig. 10 argues that the results from the P-sampled method fall significantly below this expectation. This latter conclusion is more clearly shown in Fig. 11 which shows the P-sampled variance results for $n = -1$ relative to the scale of the box. As in Eq. 25 the Gaussian variance in Fig. 11 is independent of r_0. On the largest scales, the measurements from simulations systematically fall below the Gaussian expectations while on smaller scales non-linear (i.e. higher order) contributions to the variance dominate and the comparison to pure-Gaussian statistics is no longer appropriate. The suppression of the variance of $\xi(r)$ on large scales closely resembles the prediction from Eq. 24 using a lower limit of $k = 2\pi/L_{\rm box}$ (solid gray line).

Note that [57] found good agreement comparing Eq. 25 to the covariance of the halo-halo correlation function. I point out that their comparison was made for $r/L_{\rm box} \lesssim 0.1$, where my results match the gaussian expectations without a low k cutoff.[7]

B. Var P(k) results

As discussed in the previous section the ξ-sampled method features large box-to-box variance in the correlation function. Fig. 12 shows the box-to-box variance of the power spectrum results in my simulation set. Unsurprisingly, the variance of the ξ-sampled results is much greater than the variance of the P-sampled results, which lie closer to the black-dashed line indicating the gaussian uncertainty from the number of modes at each k, i.e., $\mathrm{Var}^{1/2}(P(k))/P(k) = 2N_k^{-1/2}$. Overlaid for the P-sampled simulations in dot-dashed lines is the expected variance from non-linear effects related to the amplitude of the trispectrum [52], the fourier-space analogy of the

[7] n.b. [57], focusing on halo statistics, required careful bin-averaging of the $\sin(kr)/kr$ terms in Eq. 25 to achieve this agreement. In my matter correlation function results the bin widths are sufficiently small that this effect is negligible.

"HEPT floor" in Fig. 10. This term depends on perturbation theory results which are not expected to be extremely accurate in practice. Nevertheless the $n = -1$ results agree reasonably well, while the $n = -1.5$ and $n = -2$ cases are much less convincing.

Though this source of variance should still be a contribution to the variance in ξ-sampled simulations, it is nevertheless small and I omit this term in the lower panels of Fig. 12 for clarity, showing instead an estimate of the "beat coupling" expectation from [54]. I apply this term in much the same way as [18], directly using the variance of the DC modes in the simulations rather than $P_L(2k_{\rm box})$ to estimate this effect which is only relevant to simulations being used to produce mock catalogues. [18] find that the measured amplitude of the beat coupling term is $\sim 15\%$ below the expectation from HEPT. Compared with my results, the simulation measurements are closer to $\sim 50\%$ lower variance than this expectation. To make the most accurate mock catalogues possible, with accurate covariances, this effect should be studied more carefully. I will return to this issue in future work. The interested reader should consult [58] and references therein for a recent consideration of these issues.

C. Var hmf results

For predicting the abundance of galaxy clusters or for more general use, high precision predictions of the halo mass function are becoming increasingly important for inferring the original, linear density field (containing the cosmological information) from observations of halo number counts. Here again, there exists an expectation of the variance from a sample volume which can be compared to the simulation results. Derived originally by [59], the variance in the halo mass function should be simply related to the Poisson noise and a sample variance term,

$$\sigma_h^2 = \frac{\langle n^2 \rangle - \bar{n}_h^2}{\bar{n}_h^2} = \frac{1}{\bar{n}_h V} + b^2 \sigma^2(R_{\rm box}) \qquad (31)$$

where $\sigma(R_{\rm box})$ is the rms variance in the matter overdensity in the sample volume. The first term in Eq. 31 is the Poisson error from number counts. Separating these two sources of error in Fig. 13 I plot separately the Poisson noise with dotted lines and the expected sample variance with a black dot-dashed line. The y-axis is normalized by $\sigma(R_{\rm box})$ so that the "sample variance" curves in this plot is simply equivalent to the bias, $b(M/M_*)$. I assume the Bagla et al. [46] model for the bias as discussed in the previous section.

The measured variance in P-sampled simulations seems to fall well below the sample variance estimate; the Poisson errors dominate in all regimes. This has been noticed before in a few ΛCDM studies [60, 61] The measured variance in the ξ-sampled simulations do seem to give different predictions for the variance, although, confusingly, they do drop significantly below the expected sample variance for low M/M_*. It is not entirely clear why this is and I defer the topic to future work.

VII. SIMULATING EXTREMELY INFRARED SPECTRA

In a suite of tests I demonstrated the accuracy of the ξ-sampled method for the ensemble-averaged statistics of $\xi(r)$, $P(k)$, the halo mass function and the halo clustering bias, finding good agreement with the expected self-similar scaling of these statistics and agreement with results from P-sampled ensembles. Important differences emerged, however, in comparing the box-to-box variances of those quantities, with the ξ-sampled method predicting significantly larger variance. For the Var $\xi(r)$ results this variance greatly exceeds the expectations from Gaussian statistics and an estimate for the contribution to the variance from higher-order correlations. Deferring, for now, the question of modifying the ξ-sampled method, this result implies that this approach is substantially suboptimal for calculating the mean $\xi(r)$ and $P(k)$ in most situations of interest.

A possible exception to this, noting the remarkably robust self-similar behavior of $P(k)$ for $n = -1.5$ and -2, is in cases where one is interested in extremely red power spectra ($n_{\rm eff} \lesssim -2$) . This is a particularly difficult to simulate regime because the large-scale clustering level rapidly grows to the scale of the box, causing deviations from the true non-linear behavior. Physically, this regime is most relevant to small box simulations for modeling the "end" of the CDM hierarchy [43, 44] or (if the N-body method is extended to include hydrodynamic effects) small box cosmological simulations of galaxy formation.

To test whether the ξ-sampled method can successfully model the self-similar behavior of a power spectrum redder than $n = -2$, I ran tests with an initial $n = -2.25$ power spectrum using both ξ-sampled and P-sampled methods and similar numerical parameters to the simulations presented elsewhere in this paper. For comparison, [38] present P-sampled simulations of $n = -2.25$ with much higher resolution.

A number of issues – both technical and fundamental – make this a challenging problem for the ξ-sampled method. In the $n = -2.25$ case there is no convenient analytic or special-function form of Eq. 2 to aid in generating the ICs and, instead, the large dynamic range of the oscillations in $P_{\rm real}(k)$ must be captured entirely by numerical integration. Inaccuracies in numerical integration can also shift the distribution of Δ_i values slightly away from gaussian, an issue that can be fixed by generating the ICs for each realization (i.e. random phases and Δ_i values for the entire ensemble) *before* evolving the particle trajectories and checking first to ensure (or otherwise choosing by fiat) that $\langle \Delta_i \rangle \approx 0$ and $\langle \Delta_i^2 \rangle \approx P_{\rm real}(0)/L_{\rm box}^3$. For the first constraint, since Δ_i is a gaussian variable, the error on $\langle \Delta_i \rangle$ is simply related

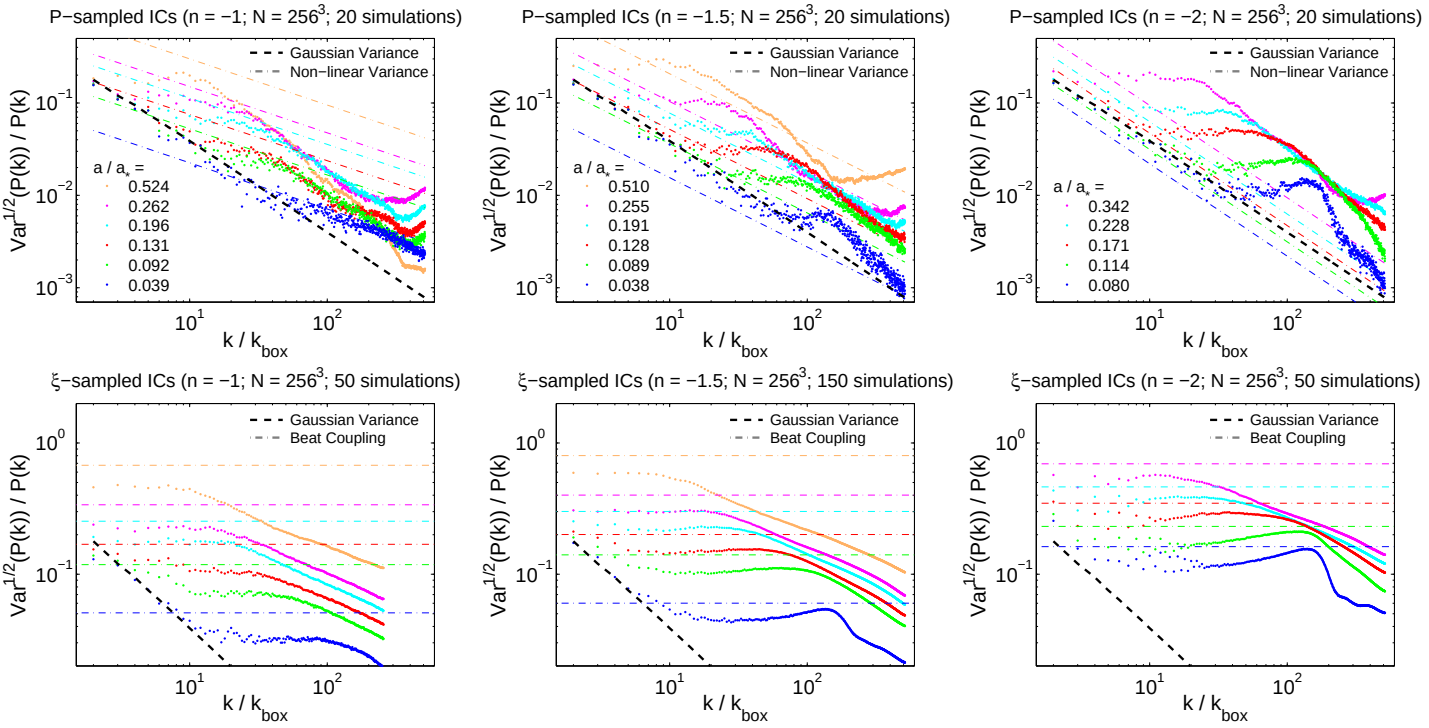

FIG. 12. Results for the box-to-box variance of $P(k)$ in my fiducial simulation ensembles. In each plot I show in black dashed lines the expected variance from mode counting ($= 2N_k^{-1/2}$). For the P-sampled simulations in the upper three panels I additionally show with colored dot-dashed lines the expected contribution to the variance from higher-order correlations [52], while for ξ-sampled simulations in the lower three panels (which, with fluctuating DC modes, are more analogous to surveys), I show expectations from the "beat coupling" result of [54] and omit the expected variance from non-linear effects for clarity.

to the variance, so $|\langle\Delta_i\rangle| < 0.01$ places the constraint,

$$N_{\rm sims} > \frac{\sqrt{\langle\Delta_i^2\rangle}}{0.01} = 100\sqrt{\frac{P_{\rm real}(0)}{L_{\rm box}^3}}. \qquad (32)$$

Notably the earlier $n = -1.5$ results did not initially meet this criterion and many more realizations were required while the 50, $n = -1$ simulations presented here do not meet this constraint, which causes the discrepancy noted previously at large $r/L_{\rm box}$ in the lower left-hand panel of Fig. 4. A more model-independent lower limit on $N_{\rm sims}$ comes from considering that there must be sufficiently many realizations to ensure that the fractional error on $\langle\Delta_i^2\rangle$ is small, i.e.,

$$\langle\Delta_i^2\rangle \gg \frac{{\rm Var}^{1/2}(\langle\Delta_i^2\rangle)}{N_{\rm sims}} = \frac{\sqrt{3\langle\Delta_i^2\rangle^2}}{N_{\rm sims}}. \qquad (33)$$

For a maximum fractional error of 4% this implies $N_{\rm sims} > 25\sqrt{3} \approx 43$, which is comparable to the number of ξ-sampled realizations in the tests reported in this paper.

Any ensemble of ξ-sampled simulations should follow the guidelines just mentioned. A peculiarity of extremely red initial power spectra is that Eq. 12 can give $P_{\rm real}(k) < 0$ at certain k values if $n < -2$. This means that the "goal" correlation function does not correspond to a physically realizable density field. In the $n = -2.25$ results shown here, I set $P_{\rm real}(k) = 0$ whenever negative $P_{\rm real}(k)$ values are reported. The consequences of this choice will be discussed shortly.

Fig. 14 compares results from 20 P-sampled simulations of $n = -2.25$ to 45 ξ-sampled simulations that have been randomly selected to ensure that the distribution of Δ_i values is sufficiently gaussian. Also shown for comparison is the non-linear fitting function for $n = -2.25$ from one $N = 1586^3$ P-sampled simulation presented in [38][8] This fitting function should be more accurate than either the P-sampled or ξ-sampled simulations which were performed with far fewer particles ($N = 256^3$). Note that, unlike the plots shown in Fig. 6, the measured error on the mean is shown for each output in the P-sampled case and for the ξ-sampled results the measured error on the mean is shown for the last output; error bars from the earlier outputs are omitted for for clarity.

Comparing the last outputs in each case, the most striking result is that at high clustering levels (large a/a_*) the P-sampled approach strongly deviates from the locus

[8] The paper contains a typo in the non-linear fitting function for $n = -2.25$ cited in Table 2. β should be 1.068 instead of 0.658 (L. Widrow private communication).

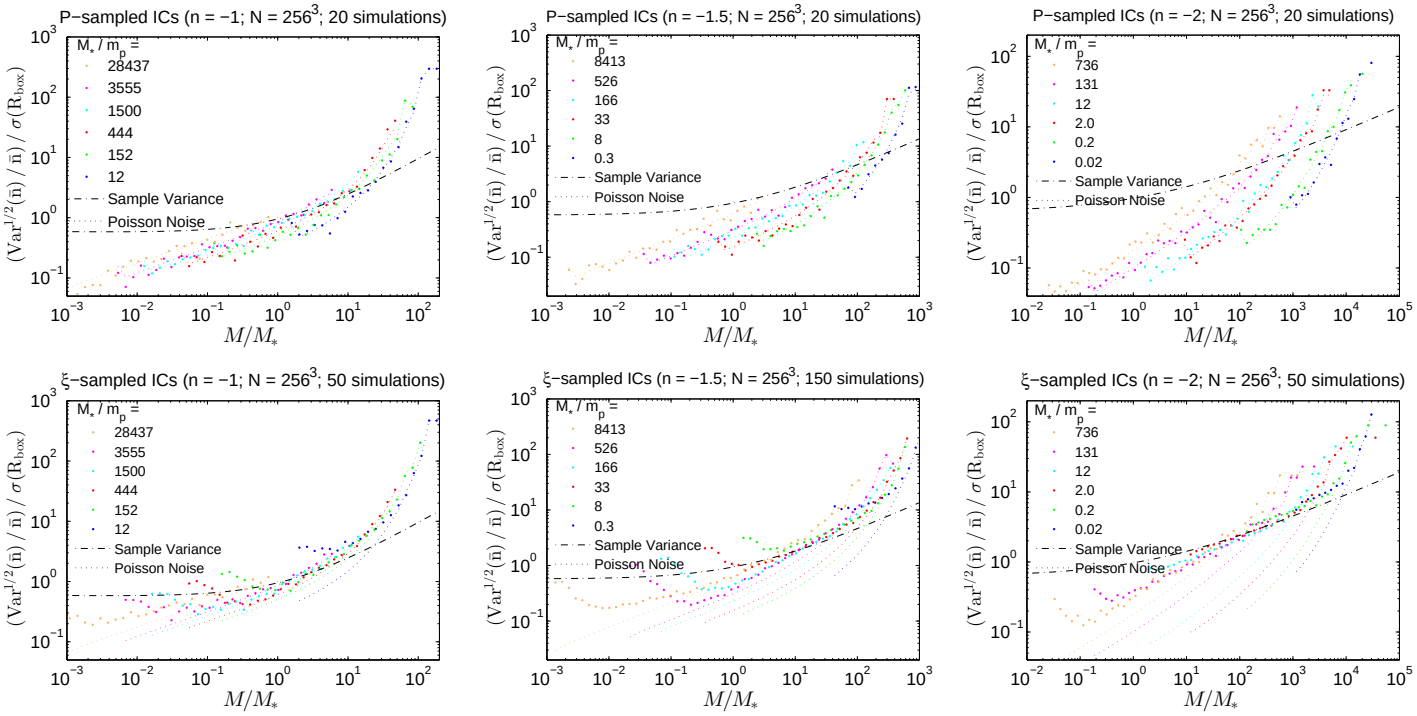

FIG. 13. Measurements of the box-to-box variance of the halo mass function in our fiducial simulation set. As in Fig. 8, the legend shows outputs in terms of the characteristic non-linear mass, M_*, and the color scheme of the outputs is consistent with other figures. The y-axis is scaled by $\sigma(R_{\rm box}) = \delta_c(R_{\rm box}/R_*)^{-(n+3)/2}$ so that the sample variance expectation (in Eq. 31) is a fixed line in each plot. Colored dotted lines show the expected variance from Poisson noise (i.e. uncertainties from counting the number of clusters).

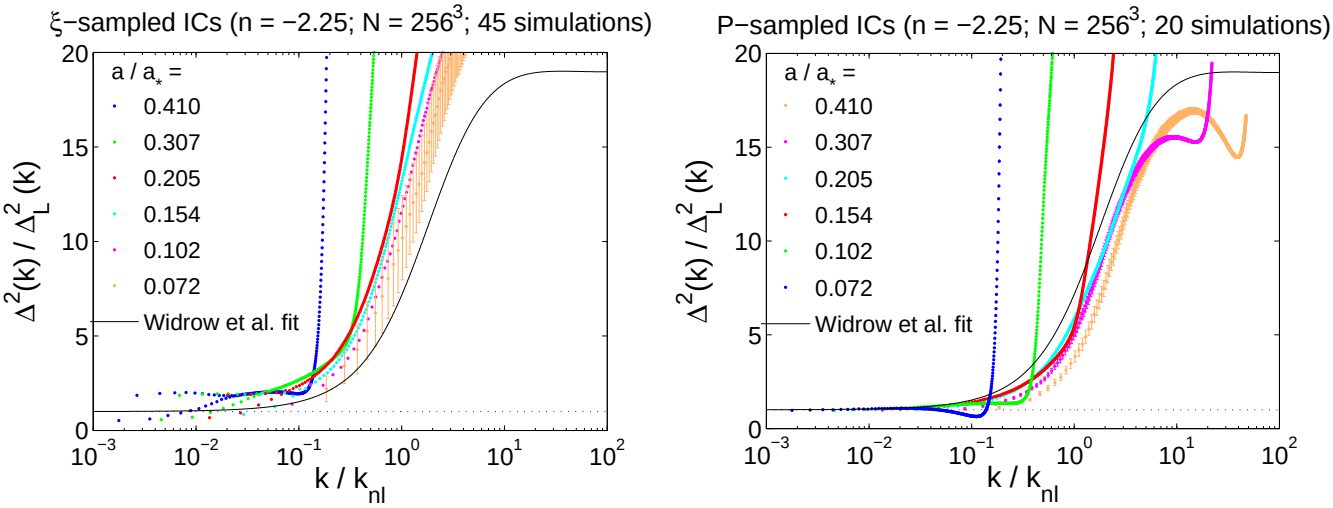

FIG. 14. Comparison of ξ-sampled (left) and P-sampled (right) simulations of an extremely red, initially $n = -2.25$ power spectrum. As in Fig. 6 the x-axis is scaled by the non-linear wavenumber, $k_{\rm nl}$, and solid black lines show results from high-resolution simulations (see text). Error bars indicating the measured error on the mean for each point are shown for all outputs in the P-sampled case while for the ξ-sampled case only the measured error on the mean for the last output is shown for clarity.

of self-similar behavior whereas the last output of the ξ-sampled case is still roughly consistent with the results from earlier outputs.

The offset between the high-resolution fitting function and the ξ-sampled results in Fig. 6 seems to be related to the zeroing of $P_{\rm real}(k)$ over certain ranges which was required to make a consistent set of initial conditions. Fig. 15 presents the results of numerically fourier transforming $P_{\rm real}(k)$ into a correlation function, $\xi_{\rm FTreal}(r)$ (Eq 1), for $n = -2.25$. The zeroing of $P_{\rm real}(k)$ causes

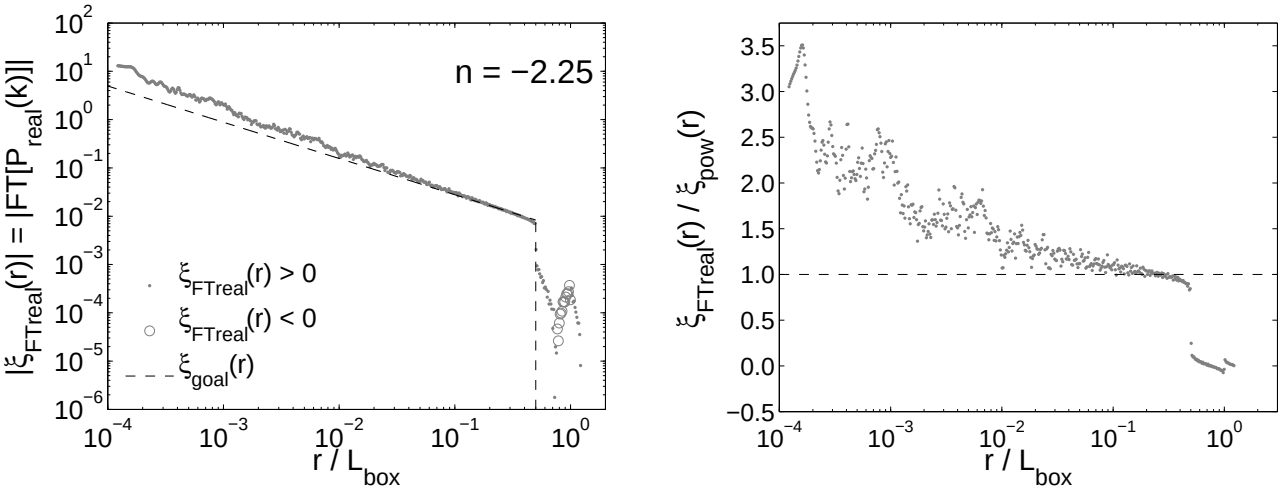

FIG. 15. Results from numerically fourier transforming $P_{\rm real}(k)$ into a correlation function, $\xi_{\rm FTreal}(r)$ (Eq. 1). In the left panel $\xi_{\rm FTreal}(r)$ (gray points and circles) is compared to the "goal" $n = -2.25$ correlation function (black dashed line). The right hand panel divides $\xi_{\rm FTreal}(r)$ by the linear theory powerlaw for $n = -2.25$.

$\xi_{\rm FTreal}(r) \neq 0$ for $r > L_{\rm box}/2$ and $\xi_{\rm FTreal}(r) \gtrsim \xi_L(r)$ for $r < L_{\rm box}/2$. Interestingly, the ξ-sampled simulation results can be brought into agreement with the high-resolution results from [38] if one concludes that the ξ-sampled simulations were initialized with a linear theory amplitude that was about 10% too high. Given the deviation of $\xi_{\rm FTreal}(r)$ from the linear theory powerlaw *above* the "goal" correlation function, shown in the right panel of Fig. 15, this seems like a plausible argument – namely that the zeroing of $P_{\rm real}(k)$ can be compensated, to first order, by a shift in the amplitude of linear theory. However, in detail, the deviations of $\xi_{\rm FTreal}(r)$ from the "goal" correlation function in Fig. 15 clearly depend on scale. Therefore, to simulate extremely red power spectra with more success than was achieved by the $n = -2.25$ results in Fig. 6, one must carefully make some assumption about $\xi(r) \neq 0$ for $r > L_{\rm box}/2$ in order to keep $P_{\rm real}(k) > 0$ and maintain the most correspondence between $\xi_{\rm FTreal}(r)$ and the "goal" correlation function. For extremely red power spectra or ΛCDM initial conditions with small box sizes a ξ-sampled approach optimized for this problem would therefore seek to *minimize* $|\xi(r)|$ for $r > L_{\rm box}/2$, rather than to simply setting the function to zero.

Assuming that a robust framework for performing this minimization can be constructed and that yet steeper power spectra can likewise maintain self-similar behavior even when $a/a_* \sim 0.4$, the ξ-sampled approach appears to be the only practical way of simulating $n_{\rm eff} \approx -2.5$. Following arguments in [38], let us assume that the smallest scales accessible to a simulation is set by half the particle nyquist wavenumber,

$$k_{\rm max} = \frac{k_{\rm Ny,p}}{2} = \frac{\pi N^{1/3}}{2L_{\rm box}} = \frac{N^{1/3}}{4} k_{\rm box}. \qquad (34)$$

Noting that $k_{\rm nl}/k_{\rm box} = (a/a_*)^{-2/(n+3)}$ [38], and assuming we are interested in scales up to $k_{\rm max}/k_{\rm nl} = 20$, the number of particles required is

$$N^{1/3} = 4\left(\frac{k_{\rm max}}{k_{\rm NL}}\right)\left(\frac{k_{\rm NL}}{k_{\rm box}}\right) = 80\left(\frac{a}{a_*}\right)^{-2/(n+3)} \qquad (35)$$

For $n = -2.5$ and assuming that the P-sampled method can work reliably out to $a/a_* \approx 0.25$ this implies $N \approx 20480^3$, significantly beyond the range of current supercomputers. By contrast, if a ξ-sampled approach can successfully capture the self-similar behavior of $P(k)$ out to $a/a_* \approx 0.45$ (as in Fig. 14), the requirement relaxes to $N \approx 2000^3$, which is well within the capabilities of current state-of-the-art machines.

Unfortunately, achiving a similar feat for small box simulations using ΛCDM initial conditions would involve a number of complications related to the difficulties already mentioned for powerlaw initial conditions. For example, characterizing the ΛCDM linear correlation function with an effective powerlaw index, $\xi_{\Lambda\rm CDM}(r) = (r_o/r)^{n_{\rm eff}(r)+3}$ (analogously to the pure powerlaw case in Eq. 11), one finds that $n_{\rm eff} \lesssim -2$ on scales as large as $r \sim 2.5h^{-1}$ Mpc. This implies that for $L_{\rm box} \lesssim 2.5h^{-1}$ Mpc, much like the ξ-sampled $n = -2.25$ pure powerlaw case, the initial power spectrum from Eq. 2 is negative for certain k values. Another problem that arises for small box ΛCDM simulations is that the need to ensure that the average overdensity of the ensemble is close to zero, as expressed in Eq. 32, leads to a prohibitively large number of simulations (see [8] Fig. 4 for the scaling of $(P_{\rm real}(0)/L_{\rm box}^3)^{1/2}$ with box size in ΛCDM). These conclusions point to the inadequacy of the ξ-sampled method for general use or even for the niche application of small box simulations. The next section explores possible modifications/improvements to the method.

VIII. DISCUSSION

Although varying the DC mode in ξ-sampled simulations correctly and naturally mends the agreement with $\xi_L(r)$ on the scale of the box, this feature comes at the cost of running many more simulations to reach the same precision on the mean $\xi(r)$ as in conventional P-sampled simulations. The indication from comparing the measured box-to-box variance of $\xi(r)$ to predictions from gaussian statistics shows that this variance far exceeds these expectations and therefore the ξ-sampled method is substantially sub-optimal for predicting the mean $\xi(r)$. Nevertheless, there are still compelling reasons to opt for the ξ-sampled approach ([7, 21], VII). Can a modified ξ-sampled approach deliver the best of both ξ-sampled and P-sampled methods?

In an effort to explore this possibility, I experimented with different choices for $a_{\rm box}/a_{\rm uni}$. In particular, one can instead choose to set $a_{\rm box}$ so that instead of ensuring that the age of the universe in each box is approximately the same, as for the choice $a_{\rm box,i}/a_{\rm uni} = 1 - \Delta_i/3$, one can match the amplitude of the linear theory growth function in each box. For an $\Omega_m = 1$ cosmology with no dark energy (as in the simulations presented throughout this paper) it can be shown that the choice $a_{\rm box,i}/a_{\rm uni} = 1 - (4/7)\Delta_i$ meets this requirement to first order. Tests using this 4/7 factor, rather than the original 1/3, further increased the variance in $\xi(r)$ (with the the box size fixed). Since 4/7 is larger than 1/3, a re-derivation of the result of Eq. 30 only *increases* the variance $\xi(r)$ about the mean, thus working in the opposite direction of the desired effect. Another, less dramatic, modification to the ξ-sampled method was made where instead of a linearized formula for $a_{\rm box}$ given $a_{\rm uni}$ and Δ as in Eq. 7, $a_{\rm box}$ was chosen by ensuring exactly that the age of the universe in a given box matched the age of the universe at the epoch of interest in the "uni" cosmology, i.e., $t_{\rm box}(a_{\rm box}) \equiv t_{\rm uni}(a_{\rm uni})$. Although $a_{\rm box}$ can deviate substantially from $a_{\rm uni}$, especially at late times, this change did not substantially affect the measured box-to-box variance of $\xi(r)$.

Another avenue to modifying the code to decrease the variance would appear to be to change the assumption that $\xi(r) = 0$ for $r > L_{\rm box}/2$ and instead allow $\xi(r)$ to be non-zero for a larger fraction of the box. In [8] the choice of $L_{\rm box}/2$ was somewhat arbitrary and a closer inspection reveals that the rms overdensity in spheres of volume equal to $L^3_{\rm box}$ deviates from the variance in the DC mode in simulation boxes with the $L_{\rm box}/2$ assumption at the $\pm 10\%$ level. However, no fixed assumption for the fraction of the box scale beyond which $\xi(r) = 0$ can bring these two statistics into perfect agreement and, for some powerlaws, correcting for this would imply setting $\xi(r) = 0$ for $r > L_{\rm box}$ (i.e. well beyond the box scale). For other powerlaws the comparison to the rms overdensity implies that the DC variance is slightly *under*-predicted. Clearly the $L_{\rm box}/2$ assumption is close enough to the true variance that an explicit correction would not uniformly move the results in the direction of more realistic box-to-box variance on the quantities of interest. Furthermore, it can be shown that changing the fraction of the box scale can not fix the the problems that arise with extremely red initial power spectra or sufficiently small boxes ($L_{\rm box} \lesssim 2.5h^{-1}$ Mpc) using ΛCDM initial conditions, which leads to negative values for $P_{\rm real}(k)$ at certain k.

Having exhausted these possibilities to improve the ξ-sampled method it would appear that one must simply abandon the need to set the variance of the DC mode exactly as $\langle \Delta_i^2 \rangle = P_{\rm real}(0)/L^3_{\rm box}$ and instead use some diminished fraction of $P_{\rm real}(0)/L^3_{\rm box}$. Essentially, the only way to retain the advantages of the P-sampled and ξ-sampled methods without dramatically increasing the number of simulations required is to allow the DC mode to to vary from simulation to simulation but not as broadly as the ξ-sampled method would prescribe. Simulations using this hybrid approach might see the suppression of $P(k)$ (Fig. 7) and the measured variance in $\xi(r)$ (Fig. 11) near the box scale but not as strongly as before. So too the agreement between the expected variance in halo number counts and the measurements from simulation might improve but not to the point of complete reconciliation. This approach would involve certain subtleties regarding the correction of the statistics of the cosmological density field described in § II C and Appendicies A & B. In particular the correlation function measurement would include a correction from both the integral constraint and the $(a_{\rm uni}/a_{\rm box})^3$ correction from the ξ-sampled method. To be sure, this hybrid method would require somewhat more simulations to reach a given precision on the mean $\xi(r)$ but only because it is being used to correct for the *suppression* of the variance of $\xi(r)$ (Fig. 11). A hybrid method along these lines would indeed have general applicability (for powerlaw ICs or ΛCDM), especially in cases when simulating more than one box to accumulate statistics is already a goal.

Although this fundamental change to the variance of the DC mode might seem poorly motivated, bear in mind that the variance in $\xi(r)$ in the lower panels of Fig. 10 was found to be dominated by the $(a_{\rm unit}/a_{\rm box})^3$ correction fator. If the variance of the DC mode was taken as a sufficiently small fraction of $P_{\rm real}(0)/L^3_{\rm box}$, the variance introduced by this correction would decrease and this could be used to mend the agreement with the gaussian statistics in Fig. 11. Introducing this parameter to suppress the variance in the DC mode simply implies a continuum between ignoring the DC mode, as in conventional P-sampled simulations, and fully including it as in the original ξ-sampled method. This would not be a free parameter in the sense that both the *suppressed* variance of $\xi(r)$ in P-sampled simulations (Eq. 24 including a cutoff at $k_{\rm box}$) and the *excess* variance in ξ-sampled simulations (Eqs. 26-29) are well understood so that the parameter would be uniquely set to reconcile as much as possible the the measured variance of $\xi(r)$ with the ex-

pectations for an infinite box. Though a very promising line of inquiry, it is beyond the scope of this paper to further investigate this hybridization of the ξ-sampled and P-sampled methods and I defer this to future work.

IX. SUMMARY

This paper presents extensive tests demonstrating the accuracy of running ensembles of cosmological N-body simulations both using conventional P-sampled ICs, as well as so-called ξ-sampled ICs, which was originally conceived by Pen [7] and implemented by Sirko [8]. Notably, I found that the conventional P-sampled results presented in [8] suffered from an integral constraint bias on the scale of the simulation box. Correcting for this *measurement* bias substantially improves the P-sampled $\xi(r)$ predictions, increasing the amplitude of $\xi(r)$ by almost a factor of two near the box scale, so that both ξ-sampled and P-sampled ensembles are unbiased estimators of the mean $\xi(r)$ and a key disadvantage to the P-sampled method is removed.

In a suite of tests I used powerlaw initial conditions (with $\Omega_m = 1, \Omega_{\rm de} = 0$) so that the scale-free nature of these models, which gives rise to self-similar evolution of the cosmological density field, could be used to assess the accuracy of the simulations on different scales and to provide an absolute reference for when for when the predictions of ξ-sampled and P-sampled ensembles disagree. Tests of this nature proved decisive in confirming the accuracy of the first generation of (P-sampled) cosmological N-body codes [22]. I focus on $n = -1, -1.5$ and -2 powerlaws as well as a BAO-like powerlaw setup introduced by [23] where powerlaws of $n = -0.5, -1$ and -1.5 are modulated by a gaussian bump feature in real space, resulting in sinusoidal oscillations around the powerlaws in fourier space. Additionally, mindful that the ξ-sampled code has been used to generate mock catalogues [18], I investigated the halo mass functions and halo clustering bias in the pure powerlaw simulations, comparing to the results of [46]. In an appendix I also compared the power spectrum measurements from the P-sampled simulations to perturbation theory predictions for the suppression of the growth of structure caused by the finite box scale, as originally proposed by [45].

Generally, in the tests presented here using $N = 256^3$ particles and the Gadget2 code, the ξ-sampled and P-sampled methods agreed on the mean $\xi(r)$, $P(k)$, halo mass function and halo clustering bias. ξ-sampled simulations of the BAO-like setup also agreed well with the $\xi(r)$ measurements from simulations presented in [23], indicating that the shift and broadening of the BAO feature are accurately modeled. Insofar as the ξ-sampled and P-sampled methods are equally valid approaches to making predictions for the non-linear growth of structure, these results reassuringly imply that non-linearities are robustly predicted by both of these methods. A difference, for example, between the P and ξ-sampled predictions for the BAO shift would have been very concerning. Another interesting conclusion, with similarly broad implications, is that the measured correlation function in both P-sampled and ξ-sampled simulations with pure powerlaws matches the self-similar solution on scales below the initial mean interparticle spacing at late times, after the particles had moved significantly from their initial positions. Although emphasizing the problems introduced by aggressive force softening such as that employed here, [41] find a similar trend in ΛCDM simulations and the results shown here extend that result to a more diverse range of initial power spectra.

Significant differences between the ξ- and P-sampled methods emerged for the box-to-box variances of those quantities and the larger box-to-box variance of the ξ-sampled results required significantly more simulations ($\gtrsim 2.5$ times more; throughout I typically compare ensembles of 20 P-sampled simulations to 50 or more ξ-sampled simulations) to achieve a given precision on the mean $\xi(r)$ and other statistics. By comparing to the expected variance of $\xi(r)$ assuming gaussian fields and linear theory, the $n = -1$ results unambiguously showed that the variance of $\xi(r)$ significantly exceeds this expectation so that the ξ-sampled method is substantially sub-optimal for predicting the mean $\xi(r)$. By contrast the P-sampled method was shown to have a box-to-box variance in $\xi(r)$ that can be as much as a factor of two *smaller* than this expectation on scales approaching the box size. Cutting off the integral in the gaussian prediction at $k_{\rm box} = 2\pi/L_{\rm box}$ (Eq. 24) adequately modeled this important deficiency of the P-sampled method.

Digging deeper into the source of the large variance of $\xi(r)$ in ξ-sampled simulations, it was revealed that this variance comes predominantly from the $(a_{\rm uni}/a_{\rm box})^3$ correction term in the $\xi(r)$ measurements. This connects the number of ξ-sampled simulations needed to reach a given precision on the mean $\xi(r)$ to the requirement that the mean overdensity in all the simulation boxes, $\langle \Delta_i \rangle$, is approximately zero and that $\langle \Delta_i^2 \rangle \approx P_{\rm real}(0)/L_{\rm box}^3$, as it must be in the $N_{\rm sims} \to \infty$ limit. These issues caused the measured mean $\xi(r)$ in $n = -1$ simulations to move well away from linear theory and the self-similar solution on scales approaching the box because of an insufficient number of simulations ($N_{\rm sims} = 50$)[9].

Having completed these tests the principal advantage of the ξ-sampled method was in predicting the mean $P(k)$ when the clustering becomes comparable to the size of the box. This is an important concern when simulating extremely steep power spectra ($n_{\rm eff} \lesssim -2$) as is the case when investigating the "end" of the CDM hierarchy with N-body simulations [43, 44] or, e.g., in simulating small boxes to study galaxy formation in the early universe. As a further test of the two methods

[9] This problem does not affect the conclusions mentioned in the previous paragraph regarding the box-to-box variance of $\xi(r)$ in the $n = -1$ case

in this regime I simulated an initially $n = -2.25$ power-law spectrum, and compared to the high resolution results in [38]. The results showed some evidence that the ξ-sampled method maintained self-similarity better than simulations using the P-sampled approach but overall did not prove steeply falling power spectra to be the "killer app" for the ξ-sampled approach. This investigation also noticed that the shape of the ΛCDM correlation function is flat enough at small scales that for $L_{\rm box} \lesssim 2.5\ h^{-1}$ Mpc the method will fail because Eq. 2 would imply $P_{\rm real}(k) < 0$ for certain k values. These conclusions all pointed to the need to modify the ξ-sampled approach for more general use.

A number of modifications to the ξ-sampled approach was discussed and investigated. Changinging the relationship between the scale factor of interest in the box given the overdensity and the scale factor of interest in an asymptotically larger box only *increased* the variance of $\xi(r)$ and $P(k)$, contrary to the desired effect. Other modifications were ruled out such as changing the scale beyond which $\xi(r) \equiv 0$ from $L_{\rm box}/2$ to some other fraction of the box size.

A modification that should work, however, is to supress the variance of the DC mode in a kind of hybrid approach where the simulations are run with less-than-prescribed scatter around the desired cosmology and the correlation function is measured with corrections arising from both the P-sampled and ξ-sampled methods. The suppression of the DC mode can be straightforwardly related to a decrease in the measured box-to-box variance of $\xi(r)$ until the gaussian expectation of the variance of $\xi(r)$ is met. Since the suppression of the variance of $\xi(r)$ in a finite, P-sampled box is well understood the parameter that suppresses the variance of the DC mode is uniquely defined. (The measurement of other statistics, e.g. $P(k)$ or halo bias, would need to be considered carefully as well.) On a basic level, introducing this suppression merely proposes a method in which there is a continuum from completely neglecting the DC mode, as in P-sampled simulations, and fully including it as in the ξ-sampled method. Such a method would be generally applicable (for powerlaws or ΛCDM ICs) and might alleviate some of the known shortcomings of the conventional P-sampled approach [7, 21].

ACKNOWLEDGEMENTS

The author thanks the Ohio State University Center for Cosmology and AstroParticle Physics for its support, and David Weinberg for guidance through the project. Thanks also goes to Jeremy Tinker for insightful conversations and Ed Sirko for helpful correspondence. A special thanks to Stelios Kazantzidis (CCAPP) and the OSU astronomy department for making available compute nodes for this project, as well as the Ohio Supercomputer Center which was also a valuable resource. This project has been supported by NSF grant AST-1009505 and AST-0707985.

Appendix A: The Integral Constraint on $\xi(r)$ in P-sampled Simulations

An important but sometimes neglected subtlety in measuring the correlation function in standard, P-sampled simulations is the artificial imposition of an integral constraint on the *measured* $\xi(r)$ on the scale of the simulation volume. This effect is entirely orthogonal to the question of which estimator [27, 28, etc.] converges most rapidly to the true $\xi(r)$ in the presence of Poisson noise. The problem stems from using a finite volume with a fixed number density of particles $\bar{n}$ to estimate $\xi(r)$. Since, in a simulation box, the integral over the number density of particles separated by distance r ($\Gamma(r)$ in Eq. 8), over the simulation volume yields the number density of particles,

$$\int_{V_{\rm box}} r^2 \Gamma(r) dr = \bar{n}, \tag{A1}$$

with $\xi(r)$ in its original form in Eq. 8, this implies

$$\int_{V_{\rm box}} r^2 \xi_{\rm meas}(r) dr = 0 \tag{A2}$$

regardless of whether this is the case with the true $\xi(r)$. In finite volumes in the real universe $\bar{n}$ will vary from box-to-box and, importantly, the average of $\bar{n}^{-1}$ (which appears in Eq. 8) will deviate from the inverse of the true (i.e. asymptotically large volume) number density. In surveys a similar problem is encountered from uncertainties and biases in the average number density of objects – in both cases the problem lies with the $\bar{n}$ term in Eq. 8 rather than the $\Gamma(r)$ term.

For ΛCDM simulations using large boxes ($L_{\rm box} \gtrsim 2h^{-1}$ Gpc) the integral constraint is a minor issue, but for simulations with smaller boxes this is an important concern. Notably, [8] presented simulations with $L_{\rm box} = 50 - 100h^{-1}$ Mpc without any kind of corrections for this effect. As a remedy consistent with the more sophisticated treatments of [56] and [27], [23, Appendix B] proposed that the correction should take the form

$$\xi(r) = \xi_{\rm meas}(r) + \bar{\xi}_L(R_{\rm box}) \tag{A3}$$

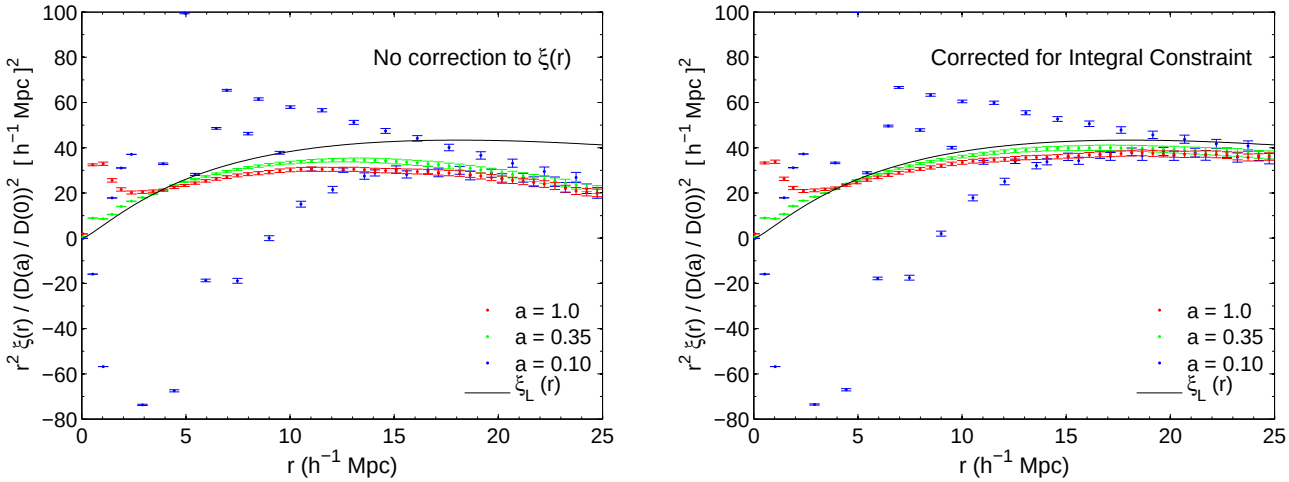

FIG. 16. Matter correlation function results from a ΛCDM ensemble of simulations ($L_{\rm box} = 100h^{-1}$ Mpc, $N = 64^3$, 100 realizations) using standard (P-sampled) ICs. The left panel shows the $\xi(r)$ results from this simulation set without applying the integral constraint correction described in the text. The right panel shows the results from including the corrections in Eq. A3. Error bars show the error on the mean. Note that the earliest output ($a = 0.1$, shown in blue) is severely affected by transients from the initial conditions.

where $\xi_{\rm meas}(r)$ is as in Eq. 8 and $\bar{\xi}_L(R_{\rm box}) = 3/R_{\rm box}^3 \int_0^{R_{\rm box}} r^2 \xi_L(r) dr$ where $\xi_L(r)$ is the linear theory correlation function and $R_{\rm box} = (4\pi/3)^{-1/3} L_{\rm box} = L_{\rm box}/1.61$.

A simple derivation can be used to estimate the bias introduced by the integral constraint for large boxes assuming a ΛCDM initial power spectrum. In this case,

$$\bar{\xi}(R_{\rm box}) = \frac{3}{4\pi R_{\rm box}^3} \int_0^{R_{\rm box}} 4\pi r^2 \xi_{\Lambda CDM}(r) dr = \frac{3}{4\pi R_{\rm box}^3} \left[\int_0^{\infty} 4\pi r^2 \xi_{\Lambda CDM}(r) dr - \int_{R_{\rm box}}^{\infty} 4\pi r^2 \xi_{\Lambda CDM}(r) dr \right]. \quad \text{(A4)}$$

The integral over infinity is eqivalent to $P(k \to 0)$ which goes to zero because $P(k) \sim k$ on large scales. The other term within the brackets can be approximated analytically since on scales larger than $r \sim 250h^{-1}$ Mpc, $\xi_{\Lambda CDM} \approx \xi_0 (r_0/r)^4$ where r_0 is a constant and the amplitude, ξ_0, is negative. It can be easily shown that

$$\bar{\xi}(R_{\rm box}) \approx 3(-\xi_0) \left(\frac{r_0}{R_{\rm box}} \right)^4 = 20.16(-\xi_0) \left(\frac{r_0}{L_{\rm box}} \right)^4. \quad \text{(A5)}$$

Applying this result to estimate the fractional bias in the amplitude of the BAO feature yields

$$\frac{\xi(r_{\rm bao}) - \hat{\xi}(r_{\rm bao})}{\xi(r_{\rm bao})} = \frac{\bar{\xi}(R_{\rm box})}{\xi(r_{\rm bao})} \approx 0.54\% \left(\frac{1h^{-1}{\rm Gpc}}{L_{\rm box}} \right)^4 \quad \text{(A6)}$$

where $\hat{\xi}(r_{\rm bao})$ is the uncorrected correlation function and I have assumed $(-\xi_0)/\xi(r_{\rm bao}) \approx$ 2.71e-4 and $r_0 \approx 1h^{-1}$ Gpc, using CAMB [62] and parameters from WMAP7 [63]. Formally, because of a cancellation of the square of the linear theory growth function in the ratio $(-\xi_0)/\xi(r_{\rm bao})$, Eq. A6 is independent of redshift and, if left unaccounted for, this measurement bias will propagate to change inferences regarding the broadening and shift of the BAO feature in the correlation function as well regardless of epoch. In more detail, redshift-dependent contributions arising from higher-order correlations can also bias the correlation function [56], so in practice Eq. A6 can be thought of as a lower bound.

Fig. 16 shows the importance of integral constraint with a set of P-sampled results from ΛCDM simulations ($\Omega_m = 0.27, \Omega_\Lambda = 0.73, h = 0.71, \sigma_8 = 0.84$) that closely mimics the fiducial simulation set in [8] ($N = 64^3$ particles, $L_{\rm box} = 100h^{-1}$ Mpc, 100 realizations with $z_i = 39$). Note that in this case the box size is small enough that the $\bar{\xi}(R_{\rm box})$ correction must be evaluated numerically.

Fig. 16 can be directly compared to Fig. 9 of [8]. Although the first output is severely affected by transients from the initial conditions I include it for comparison with [8]. The integral-constraint correction increases the correlation function at 1/4th scale of the box scale by almost a factor of two, considerably improving the agreement with linear theory. Without including this correction it is tempting to conclude incorrectly that the standard P-sampled method dramatically fails in modeling the correlation function unless box sizes much larger than the scales of interest are

used. I apply this correction to all of the P-sampled results shown in this paper. ξ-sampled results do not require this correction, and Eq. 9 naturally takes this effect into account. In fact, the correction to $\xi(r)$ in Eq. A3 bears a close resemblance to the form of Eq. 9, as I discuss in § VI.

Appendix B: The Integral Constraint in Halo Bias Measurements

As discussed in Appendix A, correlation function measurements in P-sampled simulations suffer from an integral constraint imposed on the scale of the box. Since this measurement bias arises from incorrectly assuming that the density of pairs in the simulation volume is the same as the density of pairs as in an infinite volume, these conclusions extend to clustering measurements of any distribution of points in a finite volume (e.g. dark matter particles, halos, quasars). The purpose of this appendix is to examine this subtlety for the task of computing the halo clustering bias from a set of P-sampled simulations.

For bias measurements I am interested in the quantity,

$$b^2 = \frac{\xi_{hh}(r)}{\xi_{mm}(r)} \ , \tag{B1}$$

where $\xi_{hh}(r)$ is the halo-halo correlation function and $\xi_{mm}(r)$ is the matter-matter correlation function, elsewhere referred to in this paper as $\xi(r)$. Using the corrected values of $\xi_{hh}(r)$ and $\xi_{mm}(r)$ yields,

$$b^2 = \frac{\hat{\xi}_{hh}(r) + b^2\bar{\xi}(R_{\rm box})}{\hat{\xi}_{mm}(r) + \bar{\xi}(R_{\rm box})} \tag{B2}$$

where $\hat{\xi}_{hh}(r)$ are $\hat{\xi}_{mm}(r)$ are uncorrected correlation function measurements. Moving the denominator of Eq. B2 to the left hand side,

$$b^2(\hat{\xi}_{mm}(r) + \bar{\xi}(R_{\rm box})) = \hat{\xi}_{hh}(r) + b^2\bar{\xi}(R_{\rm box}) \tag{B3}$$

the $\bar{\xi}(R_{\rm box})$ terms cancel, leaving just

$$b^2 = \frac{\hat{\xi}_{hh}(r)}{\hat{\xi}_{mm}(r)} \ . \tag{B4}$$

It is unclear whether this subtlety was noticed in previous studies where bias measurements from pure powerlaw simulations using real-space statistics have been made [36, 47]. In any case, this appendix argues that the correction cancels, at least to first order. More sophisticated models exist for correcting for the integral constraint (e.g. [17]) and obtaining unbiased estimators of the halo clustering bias. The bias results presented in Fig. 9 were determined using the simple formula of Eq. B4.

Appendix C: Modeling of the Effect of the Box Scale on the Growth of Structure in Simulations

Seto [45] provides arguments that the *integral expressions* for standard 1-loop perturbation theory (PT) [64] can be used with a truncated initial power spectrum to capture the effect of the box scale on the growth of structure in cosmological N-body simulations. Formally, then, if

$$P_{L,\rm trunc}(k) = \begin{cases} 0 & k < k_{\rm box} \\ Ak^n & k_{\rm box} < k < k_c \\ 0 & k > k_c \end{cases} \tag{C1}$$

represents the initial power spectrum that is actually represented in the simulation, then the predicted non-linear power spectrum for $k \gtrsim k_{\rm box}$ is

$$P_{1-{\rm loop},\rm trunc}(k) = P_L(k) + P_{2,\rm trunc}(k) \tag{C2}$$

where

$$P_{2,\rm trunc}(k) = P_{22,\rm trunc}(k) + P_{13,\rm trunc}(k) \tag{C3}$$

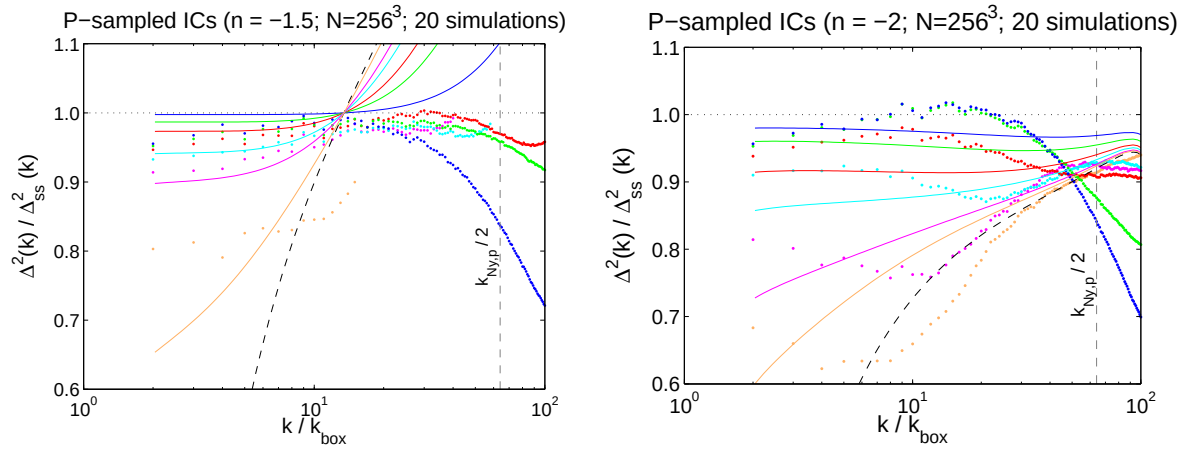

FIG. 17. A zoomed-in plot of small $k/k_{\rm box}$ results from P-sampled simulations with $n = -1.5$ (left panel) and $n = -2$ (right panel) originally presented in Fig. 7. Solid lines show an expectation for the deviations from self-similar behavior caused by the suppression of non-linear growth from the finite scale of the simulation box (Eq. C11; inspired by [45]). Colors correspond to outputs at different epochs, as in Fig. 6 and 7, with blue corresponding to the first output in each case ($a/a_* = 0.038$ for $n = -1.5$ and $a/a_* = 0.080$ for $n = -2$) and orange corresponding to the last output ($a/a_* \sim 0.5$ for both powerlaws).

and

$$P_{22,\rm trunc}(k) = \frac{k^3}{98\,(2\pi)^2}\int_0^\infty dr P_{L,\rm trunc}\,(kr)\int_{-1}^{1} dx P_{L,\rm trunc}\left[k\left(1+r^2-2rx\right)^{1/2}\right]\frac{\left(3r+7x-10rx^2\right)^2}{\left(1+r^2-2rx\right)^2} \tag{C4}$$

$$P_{13,\rm trunc}(k) = \frac{k^3 P_{L,\rm trunc}(k)}{252\,(2\pi)^2}\int_0^\infty dr P_{L,\rm trunc}(kr)\left[\frac{12}{r^2}-158+100r^2-42r^4+\frac{3}{r^3}\left(r^2-1\right)^3\left(7r^2+2\right)\ln\left|\frac{1+r}{1-r}\right|\right]. \tag{C5}$$

[45] shows the ratio $P_{2,\rm trunc}(k)/P_2(k)$ (i.e. with and without truncation) versus $k/k_{\rm box}$ for $n = 1, 0, -1$ and -2 as an expectation for how much the box scale can change the non-linear growth.

To compare with Fig. 7, although in principle $P_{1-\rm loop,trunc}(k)/P_{ss}(k) = \Delta^2_{1-\rm loop,trunc}(k)/\Delta^2_{ss}(k)$ is the correct quantity to display, since the large k predictions of 1-loop PT are typically not very accurate, it seems more meaningful to plot $P_{1-\rm loop,trunc}(k)/P_{1-\rm loop}(k)$ which, at an early enough epoch, automatically converges to one (i.e. no modification of clustering caused by the box scale) since both numerator and denominator asymptote to $P_L(k)$. For the denominator the analytic results from Appendix B of [65] are quite helpful, eliminating the need for numerical integration. In terms of the dimensionless power spectrum, [65] (see also [66]) is of the form

$$\Delta^2_{1-\rm loop}(k) = \Delta^2_L(k) + \Delta^2_2(k) = \Delta^2_L(k)\left(1+\alpha_n\Delta^2_L(k)\right) \tag{C6}$$

where α_n is a constant and

$$\Delta^2_L(k) = (k/k_{\rm nl})^{n+3} \tag{C7}$$

and

$$\Delta^2_2(k) = \alpha_n (k/k_{\rm nl})^{2(n+3)}. \tag{C8}$$

Evaluating the expressions for α_n in [65] Appendix B gives $\alpha_{-1.5} = 0.33527$ and $\alpha_{-2} = 1.56254$. The $n = -1$ powerlaw, which is UV divergent in $\Delta^2_{1-\rm loop}$, is omitted from the results presented here since, even for quasi-reasonable assumptions for k_c, the perturbation theory prediction for $\Delta^2_{1-\rm loop}(k)$ so poorly resembles the true non-linear behavior that [45] found that the non-linear clustering is *enhanced* near the box scale (since $P_{2,\rm trunc}(k)/P_2(k) > 1$ for all k), rather than suppressed as seen in Fig. 7 or, e.g, [38].

For $-3 < n < -1$, then,

$$\frac{\Delta^2_{1-\mathrm{loop,trunc}}(k)}{\Delta^2_{1-\mathrm{loop}}(k)} = \frac{\Delta^2_L(k) + \Delta^2_{2,\mathrm{trunc}}(k)}{\Delta^2_L(k) + \Delta^2_2(k)} \tag{C9}$$

$$= \frac{\Delta^2_L(k)/\Delta^2_2(k) + \Delta^2_{2,\mathrm{trunc}}(k)/\Delta^2_2(k)}{1 + \Delta^2_L(k)/\Delta^2_2(k)} \tag{C10}$$

$$= \frac{\alpha_n^{-1}(k/k_{\mathrm{nl}})^{-(n+3)} + \Delta^2_{2,\mathrm{trunc}}(k)/\Delta^2_2(k)}{1 + \alpha_n^{-1}(k/k_{\mathrm{nl}})^{-(n+3)}}. \tag{C11}$$

The term, $\Delta^2_{2,\mathrm{trunc}}(k)/\Delta^2_2(k)$, is independent of epoch and is identical to the quantity plotted in [45] Fig. 3. Fig. 17 shows the results from evaluating the numerator of this term numerically (as in [45]) while using Eqs. C8 for $\Delta^2_2(k)$. For reference, dashed black lines in Fig. 17 show the maximally allowed clustering power, which corresponds to the $k_{\mathrm{nl}} \to 0$ limit where $\Delta^2_{1-loop,\mathrm{trunc}}(k)/\Delta^2_{1-\mathrm{loop}}(k) \to \Delta^2_{2,\mathrm{trunc}}(k)/\Delta^2_2(k)$.

Qualitatively, Eq. C11 does bear some resemblance to the simulation results for $k \lesssim 10k_{\mathrm{box}}$. Note that for later outputs, in the deeply non-linear regime, one can hardly expect perfect agreement and that the non-linear fitting function for $n = -2$ is based on only one realization from [38], so the disagreement between Eq. C11 and, for example, the red ($a/a_* = 0.191$) output may be due to poor constraints on $\Delta^2_{ss}(k)$ for $k/k_{\mathrm{nl}} \ll 1$. (By contrast, the $n = -1.5$ fitting function comes from 21 realizations [23] and is unlikely to be substantially different from what is assumed here.) With or without these caveats, clearly there is some promise in using PT schemes to anticipate how accurate a given simulation may be on scales approaching the box size. And, potentially, using one of the many other, more accurate PT schemes in the literature (see, e.g., [67] and references therein) the comparison to the measured suppression of non-linear clustering power from simulation could easily be improved. These schemes should excel at robustly predicting $P(k)$ at *low* k. Typically, the concern in the literature is in the opposite regime of characterizing non-linearities at large k.

[1] J. Annis, F. J. Castander, A. E. Evrard, J. A. Frieman, E. Gaztanaga, B. Jain, A. V. Kravtsov, O. Lahav, H. Lin, J. Mohr, et al. (2005), arXiv:astro-ph/0510194.
[2] H. Wu, A. R. Zentner, and R. H. Wechsler, ApJ **713**, 856 (2010), 0910.3668.
[3] H. Seo, J. Eckel, D. J. Eisenstein, K. Mehta, M. Metchnik, N. Padmanabhan, P. Pinto, R. Takahashi, M. White, and X. Xu, ApJ **720**, 1650 (2010).
[4] K. Heitmann, M. White, C. Wagner, S. Habib, and D. Higdon, ApJ **715**, 104 (2010).
[5] S. Habib, K. Heitmann, D. Higdon, C. Nakhleh, and B. Williams, PRD **76**, 083503 (2007), arXiv:astro-ph/0702348.
[6] R. Takahashi, N. Yoshida, M. Takada, T. Matsubara, N. Sugiyama, I. Kayo, A. J. Nishizawa, T. Nishimichi, S. Saito, and A. Taruya, ApJ **700**, 479 (2009).
[7] U. Pen, ApJL **490**, L127+ (1997), arXiv:astro-ph/9709261.
[8] E. Sirko, ApJ **634**, 728 (2005).
[9] C. S. Frenk, S. D. M. White, M. Davis, and G. Efstathiou, ApJ **327**, 507 (1988).
[10] G. Tormen and E. Bertschinger, ApJ **472**, 14 (1996), arXiv:astro-ph/9512131.
[11] S. Cole, MNRAS **286**, 38 (1997), arXiv:astro-ph/9604046.
[12] R. E. Angulo and S. D. M. White, MNRAS **405**, 143 (2010), 0912.4277.
[13] M. D. Schneider, S. Cole, C. S. Frenk, and I. Szapudi, ApJ **737**, 11 (2011), 1103.2767.
[14] Y. B. Zeldovich, A&A **5**, 84 (1970).
[15] F. R. Bouchet, S. Colombi, E. Hivon, and R. Juszkiewicz, A&A **296**, 575 (1995), arXiv:astro-ph/9406013.
[16] R. Scoccimarro, MNRAS **299**, 1097 (1998), arXiv:astro-ph/9711187.
[17] J. S. Bagla and J. Prasad, MNRAS **370**, 993 (2006).
[18] B. A. Reid, D. N. Spergel, and P. Bode, ApJ **702**, 249 (2009), 0811.1025.
[19] B. A. Reid, Ph.D. thesis, Princeton University (2008).
[20] M. Tegmark, D. J. Eisenstein, M. A. Strauss, D. H. Weinberg, M. R. Blanton, J. A. Frieman, M. Fukugita, J. E. Gunn, A. J. S. Hamilton, G. R. Knapp, et al., PRD **74**, 123507 (2006), arXiv:astro-ph/0608632.
[21] N. Y. Gnedin, A. V. Kravtsov, and D. H. Rudd, ApJS **194**, 46 (2011), 1104.1428.
[22] G. Efstathiou, C. S. Frenk, S. D. M. White, and M. Davis, MNRAS **235**, 715 (1988).
[23] C. Orban and D. H. Weinberg, PRD **84**, 063501 (2011), 1101.1523.
[24] V. Springel, MNRAS **364**, 1105 (2005).
[25] Z. Lukić, K. Heitmann, S. Habib, S. Bashinsky, and P. M. Ricker, ApJ **671**, 1160 (2007), arXiv:astro-ph/0702360.
[26] M. Gabrielli, F. Sylos Labini, M. Joyce, and L. Pietronero, Statistical Physics for Cosmic Structures (Springer Press, Berlin, 2005).
[27] S. D. Landy and A. S. Szalay, ApJ **412**, 64 (1993).

[28] M. Davis and P. J. E. Peebles, ApJS **34**, 425 (1977).
[29] R. W. Hockney and J. W. Eastwood, *Computer Simulation Using Particles* (McGraw-Hill, New York, 1981).
[30] J. Einasto, A. A. Klypin, E. Saar, and S. F. Shandarin, MNRAS **206**, 529 (1984).
[31] M. Davis, G. Efstathiou, C. S. Frenk, and S. D. M. White, ApJ **292**, 371 (1985).
[32] P. J. E. Peebles, The Large-Scale Structure of the Universe (Princeton University Press, Princeton, 1980).
[33] S. N. Gurbatov, A. I. Saichev, and S. F. Shandarin, MNRAS **236**, 385 (1989).
[34] D. H. Weinberg and J. E. Gunn, MNRAS **247**, 260 (1990).
[35] A. Nusser and J. M. Colberg, MNRAS **294**, 457 (1998).
[36] Z. Zheng, J. L. Tinker, D. H. Weinberg, and A. A. Berlind, ApJ **575**, 617 (2002).
[37] R. E. Smith, J. A. Peacock, A. Jenkins, S. D. M. White, C. S. Frenk, F. R. Pearce, P. A. Thomas, G. Efstathiou, and H. M. P. Couchman, MNRAS **341**, 1311 (2003).
[38] L. M. Widrow, P. J. Elahi, R. J. Thacker, M. Richardson, and E. Scannapieco, MNRAS **397**, 1275 (2009).
[39] A. Cooray and R. Sheth, Phys. Rep. **372**, 1 (2002), arXiv:astro-ph/0206508.
[40] B. Little, D. H. Weinberg, and C. Park, MNRAS **253**, 295 (1991).
[41] M. Joyce, B. Marcos, and T. Baertschiger, MNRAS **394**, 751 (2009), 0805.1357.
[42] E. Bertschinger and J. M. Gelb, Computers in Physics **5**, 164 (1991).
[43] P. J. Elahi, R. J. Thacker, L. M. Widrow, and E. Scannapieco, MNRAS **395**, 1950 (2009), 0811.0206.
[44] D. S. Reed, R. Bower, C. S. Frenk, A. Jenkins, and T. Theuns, MNRAS **394**, 624 (2009), 0804.0004.
[45] N. Seto, ApJ **523**, 24 (1999).
[46] J. S. Bagla, N. Khandai, and G. Kulkarni, arXiv:0908.2702 (2009), 0908.2702.
[47] Y. P. Jing, ApJL **503**, L9+ (1998), arXiv:astro-ph/9805202.
[48] R. K. Sheth and G. Tormen, MNRAS **308**, 119 (1999), arXiv:astro-ph/9901122.
[49] J. L. Tinker, B. E. Robertson, A. V. Kravtsov, A. Klypin, M. S. Warren, G. Yepes, and S. Gottlöber, ApJ **724**, 878 (2010), 1001.3162.
[50] Y. P. Jing, ApJL **515**, L45 (1999), arXiv:astro-ph/9901138.
[51] A. Meiksin and M. White, MNRAS **308**, 1179 (1999), arXiv:astro-ph/9812129.
[52] R. Scoccimarro, M. Zaldarriaga, and L. Hui, ApJ **527**, 1 (1999), arXiv:astro-ph/9901099.
[53] J. D. Cohn, Nature **11**, 226 (2006), arXiv:astro-ph/0503285.
[54] A. J. S. Hamilton, C. D. Rimes, and R. Scoccimarro, MNRAS **371**, 1188 (2006), arXiv:astro-ph/0511416.
[55] R. Scoccimarro and J. A. Frieman, ApJ **520**, 35 (1999), arXiv:astro-ph/9811184.
[56] F. Bernardeau, S. Colombi, E. Gaztañaga, and R. Scoccimarro, Phys. Rep. **367**, 1 (2002), arXiv:astro-ph/0112551.
[57] A. G. Sánchez, C. M. Baugh, and R. Angulo, MNRAS **390**, 1470 (2008).
[58] R. de Putter, C. Wagner, O. Mena, L. Verde, and W. Percival, ArXiv e-prints (2011), 1111.6596.
[59] W. Hu and A. V. Kravtsov, ApJ **584**, 702 (2003), arXiv:astro-ph/0203169.
[60] M. Crocce, A. Cabré, and E. Gaztañaga, MNRAS **414**, 329 (2011), 1004.4640.
[61] S. Bhattacharya, K. Heitmann, M. White, Z. Lukić, C. Wagner, and S. Habib, ApJ **732**, 122 (2011), 1005.2239.
[62] A. Lewis, A. Challinor, and A. Lasenby, ApJ **538**, 473 (2000).
[63] E. Komatsu, K. M. Smith, J. Dunkley, C. L. Bennett, B. Gold, G. Hinshaw, N. Jarosik, D. Larson, M. R. Nolta, L. Page, et al. (2010), arXiv:1001.4538.
[64] N. Makino, M. Sasaki, and Y. Suto, PRD **46**, 585 (1992).
[65] R. Scoccimarro and J. A. Frieman, ApJ **473**, 620 (1996).
[66] R. Scoccimarro, ApJ **487**, 1 (1997).
[67] J. Carlson, M. White, and N. Padmanabhan, PRD **80**, 043531 (2009), 0905.0479.